Statistics and the Evaluation of Evidence for Forensic Scientists

WILEY SERIES IN STATISTICS IN PRACTICE

Advisory Editor, Marian Scott, University of Glasgow, Scotland, UK

Founding Editor, Vic Barnett, Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods, and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance, and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

WILEY SERIES IN STATISTICS IN PRACTICE

Advisory Editor, Marian Scott, University of Glasgow, Scotland, UK

Founding Editor, Vic Barnett, Nottingham Trent University, UK

Human and Biological Sciences

Brown and Prescott · Applied Mixed Models in Medicine

Ellenberg, Fleming and DeMets · Data Monitoring Committees in Clinical Trials: A Practical Perspective

Lawson, Browne and Vidal Rodeiro · Disease Mapping With WinBUGS and MLwiN

Lui · Statistical Estimation of Epidemiological Risk

*Marubini and Valsecchi · Analysing Survival Data from Clinical Trials and Observation Studies

Parmigiani · Modeling in Medical Decision Making: A Bayesian Approach

Senn · Cross-over Trials in Clinical Research, Second Edition

Senn · Statistical Issues in Drug Development

Spiegelhalter, Abrams and Myles · Bayesian Approaches to Clinical Trials and Health-Care Evaluation

Turner · New Drug Development: Design, Methodology, and Analysis

Whitehead · Design and Analysis of Sequential Clinical Trials, Revised Second Edition

Whitehead · Meta-Analysis of Controlled Clinical Trials

Zhou, Zhou, Liu and Ding · Applied Missing Data Analysis in the Health Sciences

Earth and Environmental Sciences

Buck, Cavanagh and Litton · Bayesian Approach to Interpreting Archaeological Data

Cooke · Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity

Gibbons, Bhaumik, and Aryal · Statistical Methods for Groundwater Monitoring, Second Edition

Glasbey and Horgan · Image Analysis in the Biological Sciences

Helsel · Nondetects and Data Analysis: Statistics for Censored Environmental Data

Helsel · Statistics for Censored Environmental Data Using Minitab® and R, Second Edition

McBride · Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions

Ofungwu · Statistical Applications for Environmental Analysis and Risk Assessment

Webster and Oliver · Geostatistics for Environmental Scientists

Industry, Commerce and Finance

Aitken and Taroni · Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition

Brandimarte · Numerical Methods in Finance and Economics: A MATLAB-Based Introduction, Second Edition

Brandimarte and Zotteri · Introduction to Distribution Logistics

Chan and Wong · Simulation Techniques in Financial Risk Management, Second Edition

Jank · Statistical Methods in eCommerce Research

Jank and Shmueli · Modeling Online Auctions

Lehtonen and Pahkinen · Practical Methods for Design and Analysis of Complex Surveys, Second Edition

Lloyd · Data Driven Business Decisions

Ohser and Mücklich · Statistical Analysis of Microstructures in Materials Science

Rausand · Risk Assessment: Theory, Methods, and Applications

Statistics and the Evaluation of Evidence for Forensic Scientists

Third Edition

Colin Aitken, Franco Taroni and Silvia Bozza

WILEY

This edition first published 2021

Edition History
John Wiley & Sons, Ltd (2e, 2004)

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Aitken, Colin, author. | Taroni, Franco, author. | Bozza, Silvia, author.
Title: Statistics and the evaluation of evidence for forensic scientists / Colin Aitken, Franco Taroni, Silvia Bozza.
Description: Third edition. | Hoboken, NJ : Wiley, [2020] | Series: Statistics in practice | Includes index.
Identifiers: LCCN 2020021384 (print) | LCCN 2020021385 (ebook) | ISBN 9781119245223 (cloth) | ISBN 9781119245254 (adobe pdf) | ISBN 9781119245414 (epub)
Subjects: LCSH: Forensic sciences – Statistical methods. | Forensic statistics. | Evidence, Expert.
Classification: LCC HV8073 .A55 2020 (print) | LCC HV8073 (ebook) | DDC 363.25/60727 – dc23
LC record available at https://lccn.loc.gov/2020021384
LC ebook record available at https://lccn.loc.gov/2020021385

Cover Design: Wiley
Cover Image: © vetalaka/Shutterstock

Set in 13/15.5pt, PhotinaMTStd by SPi Global, Chennai, India

10 9 8 7 6 5 4 3 2 1

To our families

> [T]he revision of opinion in the light of new information ... is one of the most important human intellectual activities. *(p. 290)*
>
> – *Ward Edwards*

Edwards, W. (2009). Divide and conquer: how to use likelihood and value judgments in decision making (1973). In: *A Science of Decision Making: The Legacy of Ward Edwards* (ed. J.W. Weiss and D.J. Weiss), 287–300. Oxford: Oxford University Press.

Contents

Foreword

Uncertainty affects nearly everything we do. Virtually every decision we make involves uncertainty of one kind or another. However, uncertainty does not come naturally to people's minds. Whenever we can (and sometimes when we can't), we substitute an imagined certainty that we find more comfortable and easier to plan against.

Statistics offers tools to deal with uncertainty, principally through probability. There are many models and methods in a statistician's toolkit. Which to use when, and how to create more when necessary are the typical tasks facing users of statistical methods. Every application of statistics has to be sensitive to the institutional context in which the problem arises. In the case of forensic evidence, the institutional structure includes both the organizations for which forensic scientists work and the legal structures to which they ultimately report.

The stakes are high in forensic work, as someone's liberty and/or life is typically at stake. As a

consequence, a careful consideration of the uncertainties involved is morally imperative. Doing responsible work under these circumstances requires that sources of uncertainty be identified, quantified, and reported, both truthfully and effectively. The first task is to figure out what the principal sources of uncertainty are. For example, DNA analyses often report tiny probabilities that someone other than the defendant would have the same configuration of alleles as those found at a crime scene. But this probability is premised on the assumption that the crime scene and laboratory work have been error-free. If the probability of contamination from one of these sources is one in a thousand, contamination is the dominant source of uncertainty, and should be reported.

The second task is quantification. Depending on the source of uncertainty, this can be daunting. Records can be examined to find how often collection and lab errors leading to contamination have been discovered, for example, but one is left wondering how many others there may have been that were not discovered. Experiments can help, particularly blind testing in which the technicians do not know they are being tested. Our ability to conduct such tests is in its infancy.

Finally, there is the question of how to report the uncertainty in forensic analyses. The legal structure does not necessarily welcome uncertainty, as it complicates the task of the finders-of-facts, whether judges or juries. But it is incumbent on forensic scientists to be both thoughtful and

truthful in conveying to the parties and to the court the uncertainties that lurk behind their findings. A shrill proclamation of infallibility does not advance justice.

The legal context has other implications for which statistical methods are most apt. A case involves the innocence and guilt of a particular defendant or group of defendants, faced with a particular set of evidence. As such, methods that rely for justification on long-run frequencies seem beside the point. One has to do the best one can in this specific instance. Therefore, subjective probability, which focuses on the specifics of the case without embedding it in a hypothetical infinite string of superficially similar cases, is more suited to forensic applications. What are the practical implications of such a choice? It permits forensic scientists to summarize their opinions in a number, such as 'The probability of a correspondence between the latent print at the crime scene and that of the defendant if the defendant is not the source of the crime scene print is 1%'. That's all very well as a statement of personal belief, but if anyone else is to take such a statement seriously, it must be accompanied by reasons. What assumptions were made in the analysis? What considerations make those assumptions plausible? If other plausible assumptions were made, what would their consequences be? Subjective (or personal) probability is a way of conveying one's opinion in a precise manner, but whether anyone else should pay attention to it

depends on the persuasiveness of the arguments that go with it.

The book, *Statistics and the Evaluation of Evidence for Forensic Scientists* aims to assist forensic scientists and others to do this work well. That it is now in its third edition reflects the success of the previous editions, summarizing what had been found. That a new edition is needed reflects the new thinking and new work that has been done in the last decade and a half. As more progress is made, no doubt further editions will be needed. This edition shows what has been accomplished, and charts the way forward.

December 2019 J. B. KADANE

Preface to Third Edition

In the Preface to the second edition of this book reference was made to the comment in the first edition that the role of statistics in forensic science was continuing to increase and that this was partly because of the debate continuing over DNA profiling that looked as if it would carry on into the foreseeable future. In 2004, the time of the second edition, we wrote that 'it now appears that the increase is continuing and perhaps at a greater rate than in 1995' (the time of the first edition). In 2020, we are confident that the increase is still continuing and the need for a third edition is pressing.

With the increase in the availability of data and of computing power, the role of statistical and probabilistic reasoning in the interpretation and evaluation of evidence is even more important than it was at the time of the second edition. The

courts are increasingly aware of the importance of the proper assessment of evidence in which there is random variation. Various general publications testify to the need for a new edition of this book.

- Four reports published by the Royal Statistical Society on the topic of *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice* (2010–2014) available from https://rss.org.uk/news-publication/publications/our-research/.
- Expert Evidence in Criminal Proceedings in England and Wales. The Law Commission of England and Wales, 2011, available from https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/03/lc325_Expert_Evidence_Report.pdf
- A National Academy of Sciences report *Strengthening Forensic Science in the United States: A Path Forward* (Committee on Identifying the Needs of the Forensic Sciences Community; Committee on Applied and Theoretical Statistics, National Research Council, 2009); available from https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf.
- European Network for Forensic Science Institutes Guideline for Evaluative Reporting in Forensic Science, 2015; available from http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.
- The President's Council of Advisors on Science and Technology Report on Forensic Science in

Criminal Courts: ensuring Scientific Validity of Feature-Comparison Methods, 2016; available from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

- Statistics and probability for advocates: understanding the use of statistical evidence in courts and tribunals; a report by the Inns of Court College of Advocacy and the Royal Statistical Society in 2017, available from https://rss.org.uk/news-publication/publications/our-research/.

In addition there have been two major international research programmes within the last few years. First, there was a programme on Statistics and Applied Mathematics in Forensic Science at the Statistics and Applied Mathematical Sciences Institute in North Carolina from August 2015 to May 2016; https://www.samsi.info/news-and-media/2015-16-program-on-statistics-and-applied-mathematics-in-forensic-science-forensics/. Second, there was a programme on Probability and Statistics in Forensic Science at the Isaac Newton Institute of the University of Cambridge from July to December 2016; http://www.newton.ac.uk/event/fos. There is also a Centre for Statistics and Applications in Forensic Evidence (https://forensicstats.org/), a research programme of a consortium of universities in the US which in their mission statement states that it 'brings together scientists, statisticians, forensic practitioners, and other

stakeholders to pursue the common goal of building strong statistical foundations to apply to forensic science'. The Royal Statistical Society has established a section on statistics and the law with a remit to 'improve understanding and use of statistics in the administration of justice' (https://rss.org.uk/membership/rss-groups-and-committees/sections/statistics-law/). Since 2009, the University of Lausanne (www.formation-continue-unil-epfl.ch/formation/statistics-evaluation-forensic-evidence-cas) has provided extensive on-line (e-learning) courses to train forensic practitioners in the most up to date approaches to the evaluation and interpretation of scientific evidence. The aforementioned publications and activities show the increasing interaction of law, statistics, and forensic science to try and implement societal changes in the way evidence is evaluated and interpreted.

The book is a collaboration of two statisticians and a forensic scientist. It aims to cover all material relating to an understanding of the interpretation and evaluation of evidence for trace evidence, excluding DNA evidence for which there are many specialist books. Pattern evidence and digital evidence are not covered. There are occasional mentions of fingerprint and shoeprint evidence but not in detail. The presentation is at a level that assumes only a modest mathematical and statistical background. It will help if the reader is comfortable with mathematical notation but it will be possible to skim over the more technical

parts without losing the importance of the interpretative and evaluative message. Appendices, including a detailed list of notation, will help with understanding. The ideas are illustrated with real and contemporaneous examples and data. Over 50 court cases are mentioned. There is coverage of the entire chain of reasoning with evidence from pre-assessment to presentation in court. In addition, there is consideration of measures of performance of methods of evaluation for determination of the suitability of a method for use in a particular case. The book incorporates into one volume, ideas previously discussed in separate books by the authors such as those on Bayesian networks and on Bayesian data analysis from a decision theoretic point of view. There is a comprehensive and extensive bibliography.

There has been a considerable restructuring of the book since the second edition. There has also been the introduction of new material on decision making and performance assessment. In addition, material on Bayesian networks have been introduced as part of general discussions as appropriate rather than as a separate chapter, as was the case in the second edition.

Chapter 1 on uncertainty has been expanded to include material on subjective probabilities and exchangeability. Chapter 2 on variation from the second edition has been moved to an Appendix A. This Appendix A provides source material on probability distributions for ease of reference throughout the book. Illustrative

material from the second edition has been distributed throughout the main body of the book. A few new distributions, inverse gamma, inverse chi-squared, Wishart, and inverse Wishart, have been included to cater for additional Bayesian material, with particular reference to the necessity to deal with unknown variability in univariate and multivariate Normal distributions.

The chapter on evaluation of evidence (Chapter 3 in the second edition, now Chapter 2) has been considerably increased to allow for more detailed discussion of the possible errors in interpretation and the introduction of a section on coherent decision making. The historical chapter, now Chapter 3, is little changed from the second edition; this edition is concerned with recent developments. Chapter 4 covers the Bayesian inference and incorporates material from Chapters 5 and 6 in the second edition with additional material on decision analysis and Bayesian networks.

It is with the following chapters that the greatest changes have been made. In the second edition, eighty pages were allowed for interpretation and the discussion of transfer evidence. This material is expanded to about 300 pages divided over two chapters entitled 'Evidence and propositions' and subdivided into theory (Chapter 5) and practice (Chapter 6). Bayesian networks are incorporated into these chapters. The material of Chapter 14 of the second edition is dispersed throughout the book.

Data analysis is covered in Chapter 7. The material from Chapter 9 on discrete data, Chapter 10 on continuous data, and Chapter 11 on multivariate analysis of the second edition is brought together into this one chapter for greater coherence. In the 15 years since the publication of the second edition, there has been considerable work on the assessment of performance of the models developed for the evaluation of evidence. This aspect of the topic is the subject of the final chapter, Chapter 8. The material on fibres and on DNA profiling from Chapters 12 and 13 of the second edition are dispersed throughout the book.

Reference is made on occasion to probability values of statistical distributions. We do not mention packages each time this is done, leaving the reader to use their favourite package. We have chosen the statistical package R. This is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. See https://www.r-project.org/ for further details. We make no mention of paper versions of statistical distributions, such as in books of tables, assuming that forensic scientists have access to computer systems that can access appropriate statistical software.

During the preparation of this book, several people have died to whom we owe a debt of gratitude for all that they have done for the subject and for our own careers. We remember with thanks the life and work of Annabel Bolck, Stephen Fienberg,

David Lucy, Mike Redmayne, David Schum, and Peter Tillers.

We are especially grateful to Alex Biedermann for his suggestions and helpful advice to us all throughout the preparation of this book. Others who have helped and to whom we are very grateful for their support and advice include:

Christophe Champod, Lorenzo Gaborini, Paolo Garbolino, Tacha Hicks, Graham Jackson, Agnieszka Martyna, Anders Nordgaard, Daniel Ramos-Castro, Marjan Sjerps, Amy Wilson, and Grzegorz Zadora.

Whilst we have received much advice, we accept full responsibility for any errors of commission or omission. The Leverhulme Trust, through grant number EM2016-027, provided invaluable support for this work through the award of a research fellowship to one of us (CGGA) who also thanks the School of Mathematics of the University of Edinburgh for its support. FT thanks the Swiss National Science Foundation and the Fondation pour l'Université de Lausanne for their constant support of forensic research which has permitted the development of many parts of the book. He also thanks the School of Criminal Justice of the University of Lausanne for its support. SB thanks the University Ca' Foscari of Venice for its support.

We thank Jay Kadane for agreeing to write a foreword. He has been an inspiring contributor to the subject over many years and we thank him for his support. His foreword provides a summary of the reasons why we wrote this book. We have

tried to offer solutions with explanations on how uncertainty in the evaluation and interpretation of forensic scientific evidence may be managed in the judicial process. We leave it to others to determine by how much we have succeeded.

We also thank the staff at John Wiley and Sons Ltd for their help and support in bringing this project to fruition.

Last, but by no means least, we thank our families for their support and encouragement.

C.G.G. Aitken, F. Taroni and S. Bozza
Edinburgh, Lausanne and Venice
2020

Preface to Second Edition

In the Preface to the first edition of this book, it was commented that the role of statistics in forensic science is continuing to increase and that this was partly because of the debate continuing over DNA profiling which looked as if it would carry on into the foreseeable future. It now appears that the increase is continuing and perhaps at a greater rate than nine years ago. The debate over DNA profiling continues unabated. We have left the minutiae of this debate to others restricting ourselves to an overview of that particular topic. Instead, we elaborate on the many other areas in forensic science in which statistics can play a role.

There has been a tremendous expansion in the work in forensic statistics in the nine years since the first edition of this book was published. This is reflected in the increase in the size of the book. There are about 500 pages now when there were

only about 250 in 1995, and the bibliography has increased from 10 pages to 20 pages. The number of chapters has increased from 8 to 14. The title remains the same yet there is more discussion of interpretation, in addition to new material on evaluation.

The first four chapters are on the same topics as in the first edition though the order of Chapters 2 and 3 on evaluation and on variation has been exchanged. The chapter on variation, the new Chapter 2, has been expanded to include many more probability distributions than mentioned in the first edition. As the subject has expanded so has the need for the use of more distributions. These have to be introduced sooner than before and hence the exchange of order with the chapter on evaluation. Chapter 4 has an additional section on the work of early twentieth-century forensic scientists as it has gradually emerged how far ahead of their time these scientists were in their ideas. Three new chapters have then been introduced before the chapter on transfer evidence. Bayesian inference has an increasing role to play in the evaluation of evidence yet its use is still controversial and there have been some critical comments in the courts of some of its perceived uses in the legal process. Chapter 5 provides a discussion of Bayesian inference, somewhat separate from the main thrust of the book in order to emphasise its particular relevance for evidence evaluation and interpretation. Appropriate sampling procedures are becoming ever

more important. With scarce resources, sampling provides a means of achieving almost the same inferences but with the expenditure of considerably less resources. It is important, though, that correct inferences are drawn from the results obtained from a sample. In some jurisdictions and some types of crime, such as that of drug smuggling in the United States, the quantity of illicit material associated with the crime is a factor in the sentencing. Again, if only a sample has been taken from the initial consignment, then correct inferences about the quantities involved have to be made. These are the topics of Chapter 6. Chapter 7 is a consequence of the expansion of the book to consider interpretation. This includes a discussion of work on case assessment and interpretation done since the appearance of the first edition. Chapter 7 also includes brief comments on various evidence types for which statistical evaluation is beginning to be developed. This is in contrast to those areas such as glass, fibres, and DNA profiling that are considerably more developed. Fibres and DNA each have chapters of their own, Chapters 12 and 13. Glass evaluation provides many examples throughout the book as it provides a context for much of what is discussed and it was felt this was better done at these places in the book rather than gathered together in a separate chapter. Chapters 8, 9, and 10 on transfer evidence, discrete data, and continuous data are updated versions of chapters on the same topics in the first edition. Correct analysis of multivariate data is essential

in forensic science as such data become more prevalent, for example, in the consideration of the elemental composition of glass or the chemical composition of drugs. Multivariate analysis is discussed in Chapter 11 with a worked analysis of a two-dimensional example. An appendix gives a brief description of the underlying mathematics of matrix algebra. Chapters 12 and 13 are the only chapters in the book that are specific to particular types of evidence, fibres and DNA profiling, respectively. Chapter 13 is a completely new chapter compared with the corresponding chapter in the first edition, such are the advances made in DNA profiling since the first edition appeared. It is still only a brief introduction to the topic. Other more specialised books, cited in Chapter 13, are recommended for the serious student of DNA profiling. The last chapter, 14, is an introduction to Bayesian networks, an exciting new topic for forensic science and evidence evaluation in general, predicted in the Preface to the first edition. A graphical representation, as provided by a Bayesian network, of the different types and pieces of evidence in a case aids considerably the understanding and analysis of the totality of the evidence. In addition to the bibliography and indexes, a list of notation has been provided at the end. It is hoped this will enable the reader to keep track of the symbolism which is a necessary part of ensuring clarity of exposition.

The role of Bayesian inference in forensic science remains controversial. In order to try to

understand why this may be so, we can do no better than to quote from an eminent fibres expert who wrote:

> *There may be different reasons for the obvious reluctance or scepticism connected with the adoption of Bayesian theory for presenting fibres evidence. These may include:*
>
> - *a lack of awareness of the explanatory literature available;*
> - *difficulty in understanding the proposals therein;*
> - *an antagonistic mind-set generated by an approach which is thought too complicated and too mathematical;*
> - *not knowing how to apply Bayes Theorem in practical casework;*
> - *criticism that case scenarios dealt with in the literature are over-simplified and not realistic.*
>
> *(Grieve, 2001, p. 208)*

We hope that this book goes some way towards overcoming the reluctance and scepticism.

Reference is made on occasion to probability values of statistical distributions. Rather than make reference to statistical packages and books of tables each time this is done, details of several packages are listed here instead:

MINITAB. See http://www.minitab.com and Ryan et al. (2000).

R: This is a language and environment for statistical computing and graphics which is available in source code form as Free Software under the terms of the Free Software Foundation's GNU General Public License. R can be considered as a different

implementation of S. There are some important differences, but much code written for S runs unaltered under R. See http://www.r-project.org/ and Ihaka and Gentleman (1996).

S-PLUS: See http://www.mathsoft.com/splus and Venables and Ripley (2002). See also http://lib.stat.cmu.edu/S/ for software and extensions.[1]

In addition, for those who like paper a very useful book of statistical tables is Lindley and Scott (1995).

During the preparation of this book, two eminent forensic scientists, Barry Gaudette and Mike Grieve, died. Both did much to inspire our work in evidence evaluation, for which we will always be grateful.

Many people have helped in many ways in the preparation of this book. In particular, we acknowledge the assistance of Fred Anglada, Marc Augsburger, Luc Besson, Alex Biedermann, Christophe Champod, Pierre Esseiva, Paolo Garbolino, David Lucy, Willy Mazzella, Phil Rose, and Bruce Weir. Whilst we have received much advice, we accept full responsibility for any errors of commission and omission. The Leverhulme Trust provided invaluable support of this work through the award of a research fellowship to one of us (CGGA). Dennis Lindley graciously agreed to write a foreword. He has been an inspiration to us throughout our careers and we thank him most

[1] Added in 2020: these websites no longer exist. S-Plus is owned by TIBCO (https://www.tibco.com/).

sincerely for the honour he has done us. We also thank the staff at John Wiley and Sons, Ltd., Lucy Bryan, Rob Calver, Siân Jones, Jane Shepherd, and a very assiduous copy-editor, Richard Leigh, for their help and support in bringing this project to fruition.

Last, but by no means least, we thank our families for their support and encouragement.

C.G.G. Aitken and F. Taroni
Edinburgh and Lausanne
2004

Preface to First Edition

In 1977 a paper by Dennis Lindley was published in *Biometrika* with the simple title 'A problem in forensic science'. Using an example based on the refractive indices of glass fragments, Lindley described a method for the evaluation of evidence that combined the two requirements of the forensic scientist, those of comparison and significance, into one statistic with a satisfactorily intuitive interpretation. Not unnaturally the method attracted considerable interest amongst statisticians and forensic scientists interested in seeking good ways of quantifying their evidence. Since then, the methodology and underlying ideas have been developed and extended in theory and application into many areas. These ideas, often with diverse terminology, have been scattered throughout many journals in statistics and forensic science and, with the advent of DNA profiling,

in genetics. It is one of the aims of this book to bring these scattered ideas together and, in so doing, to provide a coherent approach to the evaluation of evidence.

The evidence to be evaluated is of a particular kind, known as transfer evidence, or sometimes trace evidence. It is evidence that is transferred between the scene of a crime and a criminal. It takes the form of traces – traces of DNA, traces of blood, of glass, of fibres, of cat hairs, and so on. It is amenable to statistical analyses because data are available to assist in the assessment of variability. Assessments of other kinds of evidence, for example, eyewitness evidence, is not discussed.

The approach described in this book is based on the determination of a so-called likelihood ratio. This is a ratio of two probabilities, the probability of the evidence under two competing hypotheses. These hypotheses may be that the defendant is guilty and that he is innocent. Other hypotheses may be more suitable in certain circumstances and various of these are mentioned as appropriate throughout the book.

There are broader connections between statistics and matters forensic which could perhaps be covered by the title 'forensic statistics' and which are not covered here, except briefly. These might include the determination of a probability of guilt, both in the dicta 'innocent until proven guilty' and 'guilty beyond reasonable doubt'. Also, the role of statistical experts as expert witnesses presenting statistical assessments of data or as

consultants preparing analyses for counsel is not discussed, nor is the possible involvement of statisticians as independent court assessors. A brief review of books on these other areas in the interface of statistics and the law is given in Chapter 1. There have also been two conferences on forensic statistics (Aitken 1991, Kaye 1993) with a third to be held in Edinburgh in 1996. These have included forensic science within their programme but have extended beyond this. Papers have also been presented and discussion sessions held at other conferences, e.g. Aitken (1993) and Fienberg and Finkelstein (1996).

The role of uncertainty in forensic science is discussed in Chapter 1. The main theme of the book is that the evaluation of evidence is best achieved through consideration of the likelihood ratio. The justification for this and the derivation of the general result is given in Chapter 2. A correct understanding of variation is required in order to derive expressions for the likelihood ratio and variation is the theme for Chapter 3 where statistical models are given for both discrete and continuous data. A review of other ways of evaluating evidence is given in Chapter 4. However, no other appears, to the author at least, to have the same appeal, both mathematically and forensically as the likelihood ratio and the remainder of the book is concerned with applications of the ratio to various forensic science problems. In Chapter 5, transfer evidence is discussed with particular emphasis on the importance of the

direction of transfer, whether from the scene of the crime to the criminal or vice versa. Chapters 6 and 7 discuss examples for discrete and continuous data, respectively. The final chapter, Chapter 8, is devoted to review of DNA profiling, though, given the continuing amount of work on the subject, it is of necessity brief and almost certainly not completely up-to-date at the time of publication.

In keeping with the theme of the Series, Statistics in Practice, the book is intended for forensic scientists as well as statisticians. Forensic scientists may find some of the technical details rather too complicated. A complete understanding of these is, to a large extent, unnecessary if all that is required is an ability to implement the results. Technical details in Chapters 7 and 8 have been placed in Appendices to these Chapters so as not to interrupt the flow of the text. Statisticians may, in their turn, find some of the theory, for example, in Chapter 1, rather elementary and, if this is the case, then they should feel free to skip over this and move on to the more technical parts of the later chapters.

The role of statistics in forensic science is continuing to increase. This is partly because of the debate continuing over DNA profiling that looks as if it will carry on into the foreseeable future. The increase is also because of increasing research by forensic scientists into areas such as transfer and persistence and because of increasing numbers of data sets. Incorporation of subjective probabilities will also increase, particularly through

the role of Bayesian belief networks (Aitken and Gammerman, 1989) and knowledge-based systems (Buckleton and Walsh, 1991, Evett, 1993b).

Ian Evett and Dennis Lindley have been at the forefront of research in this area for many years. They have given me invaluable help throughout this time. Both made extremely helpful comments on earlier versions of the book for which I am grateful. I thank Hazel Easey for the assistance she gave with the production of the results in Chapter 8. I am grateful to Ian Evett also for making available the data in Table 7.3. Thanks are due to The University of Edinburgh for granting leave of absence and to my colleagues of the Department of Mathematics and Statistics in particular for shouldering the extra burdens such leave of absence by others entails. I thank also Vic Barnett, the Editor of the Series, and the staff of John Wiley and Sons Ltd. for their help throughout the gestation period of this book.

Last, but by no means least, I thank my family for their support and encouragement.

Edinburgh, 1995 C.G.G. Aitken

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

1

Uncertainty in Forensic Science

1.1 INTRODUCTION

The purpose of this book is to discuss the statistical and probabilistic evaluation of scientific evidence for forensic scientists. A formal definition of *evidence* is given in Section 1.4. For the most part the evidence to be evaluated will be the so-called *transfer* or (physical) *trace* evidence, but the general logic also applies to other types of evidence, such as digital, pattern, or testimonial evidence.

There is a well-known principle in forensic science known as *Locard's principle*, which states that every contact leaves a trace (Locard 1920).

> *[. . .] tantôt le malfaiteur a laissé sur les lieux les marques de son passage, tantôt, par une action inverse, il a emporté sur son corps ou sur ses vêtements, les indices de son séjour ou de son geste. (p. 139)*

This is translated as (Inman and Rudin 2001)

> *[. . .] either the wrong-doer has left signs at the scene of the crime, or, on the other hand, has taken away with him – on his person (body) or clothes – indications of where he has been or what he has done. (p. 93)*

The principle was reiterated using different words in Locard (1929). This has been translated by the same author in 1930. Locard (1930) wrote

> *For the microscopic debris that covers our clothes and bodies are the mute witnesses, sure and faithful, of all our movements and of all our encounters. (p. 276)*

Transfer evidence and Locard's principle may be illustrated as follows. Suppose a person gains entry to a house by breaking a window and assaults the man of the house, during which assault blood is spilt by both victim and assailant. The criminal may leave traces of their presence at the crime scene in the form of bloodstains from the assault and fibres from his clothing. This evidence is said to be transferred from the criminal to the scene of the crime. The criminal may also take traces of the crime scene away with them. These could include bloodstains from the assault victim, fibres of their clothes, and fragments of glass from the broken window. Such evidence is said to be transferred to the criminal from the crime scene. A *person of interest*[1] (PoI) is soon identified, at a time at which

[1]In earlier editions the term *suspect* was used in this context. It is now felt that the term *person of interest* is a more accurate description of the status of the person. The term *suspect* will still be used if use of 'person of interest' would be clumsy. The term *defendant* will only be used if the context is clearly that of a court.

they will not have had the opportunity to change their clothing. The forensic scientists examining the PoI's clothing find similarities amongst all the different types of evidence: blood, fibres, and glass fragments. They wish to *evaluate this evidence*. It is hoped that this book will enable them so to do.

However, for evaluation, it is not only similarity that is important but also the rarity of the characteristics of interest. Hence, quantitative issues relating to the distribution of these characteristics will be discussed. However, there will also be discussion of qualitative issues such as the choice of a suitable population against which variability in the measurements of the characteristics of interest may be compared. Also, a brief history of statistical aspects of the evaluation of evidence is given in Chapter 3.

1.2 STATISTICS AND THE LAW

The book does not focus on the use of statistics and probabilistic thinking for legal decision making, other than by occasional reference. Also, neither the role of statistical experts as expert witnesses presenting statistical assessments of data nor their role as consultants preparing analyses for counsel is discussed. There is a distinction between these two issues (Fienberg 1989, Tribe 1971). The main focus of this book is on the assessment of evidence for forensic scientists, in particular for *identification* purposes. The process of addressing the issue of whether or not a particular item

came from a particular source is most properly termed *individualization*. 'Criminalistics is the science of individualization' (at p. 236) as defined by Kirk (1963) but established forensic and judicial practices have led to it being termed *identification*. The latter terminology will be used throughout this book. An *identification*, however, is more correctly defined as 'the determination of some set to which an object belongs or the determination as to whether an object belongs to a given set' (Kingston 1965a). Further discussion is given by Kwan (1977), Evett et al. (1998), and Champod et al. (2016b). For a critical discussion of individualisation as a decision, see Cole (2014), Biedermann et al. (2008a), and Saks and Koehler (2008). More details are given in Section 2.5.9.

For example, in a case involving a broken window, similarities may be found between the refractive indices of fragments of glass found on the clothing of a PoI and the refractive indices of fragments of glass from the broken window. The assessment of this evidence, in consideration of the association or otherwise of the PoI with the scene of the crime, is part of the focus of this book.

For those interested in the issues of statistics and the law beyond those of forensic science, in the sense used in this book, there are several books available and some of these are discussed briefly.

'The evolving role of statistical assessments as evidence in the courts' is the title of a report, edited by Fienberg (1989), by the Panel on Statistical Assessments as Evidence in the Courts formed

by the Committee on National Statistics and the Committee on Research on Law Enforcement and the Administration of Justice of the United States, and funded by the National Science Foundation. Through the use of case studies, the report reviews the use of statistics in selected areas of litigation, such as employment discrimination, antitrust litigation, and environment law. One case study is concerned with identification in a criminal case. Such a matter is the concern of this book, and the ideas relevant to this case study, which involves the evidential worth of similarities amongst human head hair samples, will be discussed in greater detail later (Section 3.5.5). The report makes various recommendations, relating to the role of the expert witness, pretrial discovery, the provision of statistical resources, the role of court-appointed experts, the enhancement of the capability of the fact-finder, and statistical education for lawyers.

Two books that take the form of textbooks on statistics for lawyers are Vito and Latessa (1989) and Finkelstein and Levin (2015). The former focusses on the presentation of statistical concepts commonly used in criminal justice research. It provides criminological examples to demonstrate the calculation of basic statistics. The latter introduces rather more advanced statistical techniques and again uses case studies to illustrate the techniques.

Historically, the particular area of discrimination litigation is covered by a set of papers edited by Kaye and Aickin (1986). This starts

by outlining the legal doctrines that underlie discrimination litigation. In particular, there is a fundamental issue relating to discrimination in hiring. The definition of the relevant market from which an employer hires has to be made very clear. For example, consider the case of a man who applies, but is rejected, for a secretarial position. Is the relevant population the general population, the representation of men amongst secretaries in the local labour force, or the percentage of male applicants? The choice of a suitable reference population is also one with which the forensic scientist has to be concerned. This is discussed at several points in this book, see, for example, Sections 5.5.3.4 and 6.1.

Another textbook, which comes in two volumes, is Gastwirth (1998a,b). The book is concerned with civil cases and 'is designed to introduce statistical concepts and their proper use to lawyers and interested policy makers . . .' (volume 1, p. *xvii*). Two areas are stressed, which are usually given less emphasis in most statistical textbooks. The first area is concerned with measures of relative or comparative inequality. These are important because many legal cases are concerned with issues of fairness or equal treatment. The second area is concerned with the combination of results of several related statistical studies. This is important because existing administrative records or currently available studies often have to be used to make legal decisions and public policy; it is not possible to undertake further research. Gastwirth

(2000) has also edited a collection of essays on statistical science in the courtroom, some of which are directly relevant for this current book and will be referred to as appropriate.

A collection of papers on Statistics and Public Policy has been edited by Fairley and Mosteller (1977). One issue in the book, which relates to a particularly infamous case, the *Collins* case, is discussed in detail later (Section 3.4). Other articles concern policy issues and decision making.

Of further interest is a book (Kadane 2008) explicitly entitled 'Statistics and the Law', which considers the question 'how can lawyers and statisticians best collaborate in a court of law to present statistics in the most clear and persuasive manner?'. With the use of case studies that refer to employment, jury behaviour, fraud, taxes, and aspects of patents, there is clarification of what a statistician and what a lawyer should know for a fruitful collaboration.

Other recent publications on the interaction between law and statistics are, for example, Dawid et al. (2014), Kadane (2018a,b), Kaye (2017a,b), and Gastwirth (2017).

The remit of this book is one that is not covered by these others in great detail. The use of statistics in forensic science in general is discussed in a collection of essays edited by Aitken and Stoney (1991). The remit of this book is to describe statistical procedures for the evaluation of evidence for forensic scientists. This will be done primarily through a Bayesian approach, the principle of

which was described in Peirce (1878). It was developed further in the work of I.J. Good and A.M. Turing as code-breakers at Bletchley Park during World War Two. A brief review of the history was given in Good (1991). A history of the Bayesian approach for a lay audience was given in Bertsch McGrayne (2011). An essay on the topic of probability and the weighing of evidence was written by Good (1950). This also referred to entropy (Shannon 1948), the expected amount of information from an experiment, and Good remarked that the expected weight of evidence in favour of a hypothesis H and against its complement $\bar{H}$ is equal to the difference of the entropies assuming H and $\bar{H}$, respectively. A brief discussion of a frequentist approach and the problems associated with it is given in Section 3.6 (see also Taroni et al. 2016). A general review of the Bayesian approach was given by Fienberg (2006).

It is of interest to note that a high proportion of situations involving the so-called objective presentation of statistical evidence uses the frequentist approach with tests of significance (Fienberg and Schervish 1986).[2] However, Fienberg and Schervish go on to say that the majority of examples cited for the use of the Bayesian

[2]There is a recent critical discussion on the use of tests of significance. See, for example, Wasserstein et al. (2019), Amrhein et al. (2019), Ioannidis (2019), Haaf et al. (2019), and Johnson (2019). Note that such a discussion dates back to 1986 (Kaye 1986a) with a later update (Kaye 2017c).

approach are in the area of identification evidence. It is this area that is the main focus of this book, and it is Bayesian analyses that will form the basis for the evaluation of evidence as discussed here. Examples of the applications of such analyses to legal matters include Cullison (1969), Finkelstein and Fairley (1970, 1971), Fairley (1973), Lempert (1977), Lindley (1975, 1977b,c), Fienberg and Kadane (1983), Lempert (1986), Redmayne (1995, 1997), Friedman (1996), Redmayne (2002), Anderson et al. (2005), Robertson et al. (2016), and Adam (2016).

Another approach that will not be discussed here is that of Shafer (1976, 1982). This concerns so-called *belief functions*, see Section 3.1. The theory of belief functions is a very sophisticated theory for assessing uncertainty that endeavours to answer criticisms of both the frequentist and Bayesian approaches to inference. Belief functions are non-additive in the sense that belief in an event A [denoted Bel(A)] and belief in the opposite of A [denoted Bel($\bar{A}$)] do not sum to 1. See also Shafer (1978) for a historical discussion of non-additivity. Further discussion is beyond the scope of this book. Practical applications are few. One such, however, is to the evaluation of evidence concerning the refractive index of glass (Shafer 1982). More recent developments of the role of belief functions in law for burdens of proof and in forensic science with a discussion of the island problem (Section 6.1.6.3) and parental identification are given in Nance (2019) and Kerkvliet and Meester (2016).

It is very tempting when assessing evidence to try to determine a value for the probability of the so-called *probandum* of interest (or the ultimate issue) such as the true guilt of a PoI (as distinct from a verdict of guilty, which may or may not be correct), or a value for the odds in favour of guilt and perhaps even to reach a decision regarding the PoI's guilt. However, this is the role of the jury and/or judge. It is not the role of the forensic scientist or statistical expert witness to give an opinion on this (Evett 1983). It is permissible for the scientist to say that the evidence is 1000 times more likely, say, if the PoI is the offender than if he is not the offender. It is not permissible to interpret this to say that, because of the evidence, it is 1000 times more likely that the PoI is the offender than is not the offender. Some of the difficulties associated with assessments of probabilities are discussed by Tversky and Kahneman (1974) and are further described in Section 2.5. An appropriate representation of probabilities is useful because it fits the analytic device most used by lawyers, namely, the creation of a story. This is a narration of events 'abstracted from the evidence and arranged in a sequence to persuade the fact-finder that the story told is the most plausible account of "what really happened" that can be constructed from the evidence that has been or will be presented' (Anderson and Twining 1998, p. 166). Also of relevance is Kadane and Schum (1996), which provides a Bayesian analysis of evidence in the *Sacco–Vanzetti* case (Sacco 1969)

based on subjectively determined probabilities and assumed relationships amongst evidential events. A similar approach is presented in Section 2.9.

1.3 UNCERTAINTY IN SCIENTIFIC EVIDENCE

Scientific evidence requires considerable care in its interpretation (Evett 2009). Emphasis needs to be put on the importance of asking the question: what do the results mean in this particular case? (Jackson 2000). Kirk and Kingston (1964) emphasised:

> *Suppose that the fibres do match – what does it mean? Suppose that there is a defined degree of similarity in the bullet marking, or the handwriting, does it prove identity of origin, or does it merely give a sometimes controversial basis for making a decision as to the identity of origin? (p. 439)*

Scientists and jurists have to '[. . .] abandon the idea of absolute certainty so that a fully objective approach to the problem can be made. [. . .] If it can be accepted that nothing is absolutely certain then it becomes logical to determine the degree of confidence that may be assigned to a particular belief' (Kirk and Kingston 1964, p. 435). On the same line of reasoning, the authors (Kingston and Kirk 1964) expressed themselves on *uncertainty* and they emphasised:

> *A statistical analysis is used when uncertainty must exist. If there were a way of arriving to a* certain *answer to a*

> *problem, statistical methods would not be used. But when uncertainty does exist, and a statistical approach is possible, then this approach is the best one available since it offers an index on the uncertainty based upon a precise and logical line of reasoning. [. . .] It is undoubtedly true that serious errors have been made in applying incorrect statistical methods to the evaluation of physical evidence, but such misuse does not support the generalisation that statistics cannot be properly used in criminalistics at all. (p. 516)*

There are various kinds of problems concerned with the random variation naturally associated with scientific observations. There are problems concerned with the definition of a suitable reference population against which concepts of rarity or commonality may be assessed. There are problems concerned with the choice of a measure of the value of the evidence.

The effect of the random variation can be assessed with the appropriate use of probabilistic and statistical ideas. There is variability associated with scientific observations. Variability is a phenomenon that occurs in many places. People are of different sexes, determination of which is made at conception. People are of different height, weight, and intellectual ability, for example. The variation in height and weight is dependent on a person's sex. In general, females tend to be lighter and shorter than males. However, variation is such that there can be tall, heavy females and short, light males. At birth, it is uncertain how tall or how heavy the baby will be as an adult. However, at birth, it is usually known whether

the baby is a boy or a girl. This knowledge affects the uncertainty associated with the predictions of adult height and weight.

People are of different blood groups. A person's blood group does not depend on the age or sex of the person but does depend on the person's ethnicity. The refractive index of glass varies within and between windows. Observation of glass as to whether it is window or bottle glass will affect the uncertainty associated with the prediction of its refractive index and that of other pieces of glass, which may be thought to come from the same origin.

It may be thought that, because there is variation in scientific observations, it is not possible to make quantitative judgements regarding any comparisons between two sets of observations. The two sets are either different or they are not and there is no more to be said. However, this is not so. There are many phenomena that vary but they vary in certain specific ways. It is possible to represent these specific ways mathematically. Various probability distributions to represent variation are introduced in Appendix A. It is then possible to assess differences quantitatively and to provide a measure of uncertainty associated with such assessments.

It is useful to recognise the distinction between *statistics* and *probability*. Probability is a deductive process that argues from the general to the particular. Consider a fair coin, i.e. one in which when tossed the probability of a head landing uppermost

equals the probability of a tail landing uppermost equals 1/2. A fair coin is tossed 10 times. Probability theory enables a determination to be made of the probability that there are three heads and seven tails, say. The general concept of a fair coin is used to determine something about the outcome of the particular case in which it was tossed 10 times.

On the other hand, statistics is an inductive process that argues from the particular to the general. Consider a coin that is tossed ten times and there are three heads and seven tails. Statistics enables the question as to whether the coin is fair or not to be addressed. The particular outcome of three heads and seven tails in ten tosses is used to determine something about the general case of whether the coin was fair or not.

Fundamental to both statistics and probability is *uncertainty*. Given a fair coin, the number of heads and tails in ten tosses is uncertain. The probability associated with each outcome may be determined but the actual outcome itself cannot be predicted with certainty. Given the outcome of a particular sequence of 10 tosses, information is then available about the fairness or otherwise of the coin. For example, if the outcome were 10 heads and no tails, one may believe that the coin is double-headed but it is not certain that this is the case. There is still a non-zero probability (1/1024) that 10 tosses of a fair coin will result in 10 heads. Indeed this has occurred in the first author's own experience. A class of some 130 students were asked to each toss a coin 10 times. One student tossed 10

consecutive heads from what it is safe to assume was a fair coin. The probability of this happening is $1-(1-1/1024)^{130} \simeq 130/1024 = 0.13$. Probability is therefore the measure of choice for the quantification of uncertainty. It is therefore important to define probability carefully. This point is clarified in Section 1.7. At present, it suffices to mention a brief definition by de Finetti (1968).

> *[A probability is]* subjective *[and it] means the degree of belief (as actually held by someone, on the ground of his whole knowledge, experience, information) regarding the truth of a sentence or event, E (a fully specified single event or sentence, whose truth or falsity is, for whatever reason, unknown to that person). (p. 45)*

1.3.1 The Frequentist Method

Consider a consignment of compact disks (CDs), containing N disks. The consignment is said to be of size N. It is desired to make inferences about the proportion θ $(0 \leq \theta \leq 1)$ of the consignment which is pirated. It is not practical to inspect the whole consignment so a sample of size n, where $n < N$ is inspected.

The frequentist method assumes that the proportion θ of the consignment that is pirated is unknown but fixed. The data, that is the number of CDs in the sample that are pirated, are variable. A so-called *confidence interval* is calculated. The name *confidence* is used since no probability can be attached to the uncertain event that the interval contains θ. These ideas are discussed further in Chapter 4.

The frequentist method derives its name from the relative frequency definition of probability. The probability that a particular event, A, say, occurs is defined as the relative frequency of the number of occurrences of event A compared with the total number of occurrences of all possible events, over a long run of observations, conducted under identical conditions of all possible events. The limitations of such a definition are presented in Section 1.7.4.

For example, consider tossing a coin n times. It is not known if the coin is fair. The outcomes of the n tosses can be used as information from which the probability of a head occurring on an individual toss may be assigned. There are two possible outcomes, heads (H) and tails (T). Let $n(H)$ be the number of H and $n(T)$ be the number of T such that $n(H) + n(T) = n$. Then the probability of tossing a head on an individual toss of the coin is defined as the limit as $n \to \infty$ of the fraction $n(H)/n$. The frequentist approach relies on a belief in the long-run repetition of trials[3] under identical conditions. This is an idealised situation, seldom, if ever, realised in practice. More discussion on the interpretation of such a result is given in Section 3.6.

The way in which statistics and probability may be used to evaluate evidence is the theme of this book. Care is required. Statisticians are familiar

[3]The use of the word *trial* here is a statistical one and is not to be confused with the legal use. In a statistical use, a trial is a particular event, such as the toss of a coin. More details are in the Appendix A.2.2.

with variation, as are forensic scientists who observe it in the course of their work. Lawyers, however, prefer certainties. A defendant is found *guilty* or *not guilty* (or also, in Scotland, *not proven*). The scientist's role is to testify to the worth of the evidence, the role of the statistician and this book is to provide the scientist with a quantitative measure of this worth. It is shown that there are few forms of evidence that are so definite that statistical treatment is neither needed nor desirable. It is up to other people (the judge and/or the jury) to use this information as an aid to their deliberations. It is for neither the statistician nor the scientist to pass judgement (Kind 1994).

The use of these ideas in forensic science is best introduced through the discussion of several examples. These examples will provide a constant theme throughout the book. Consideration in detail of populations from which the criminal may be thought to have come, to which reference is made in the following text, are discussed in Section 6.1.1 where they are called *relevant populations*. The value of evidence is measured by a statistic known as the *likelihood ratio* and its logarithm. These are introduced in Sections 2.3 and 2.4.

1.3.2 Stains of Body Fluids

Example 1.1. A crime is committed. A blood-stain is found at the scene of the crime. All innocent explanations for the presence of the stain are eliminated. A PoI is found. Their DNA profile

is established and found to correspond to that of the crime stain. What is the evidential value of this correspondence? This is a very common situation yet the answer to the question provides plenty of opportunity for discussion of the theme of this book.

Certain other questions need to be addressed before this particular one can be answered. Where was the crime committed, for example? Does it matter? Does the value of the evidence of the bloodstain change depending on where the crime was committed?

Apart from their DNA profile, what else is known about the criminal? In particular, is there any information, such as ethnicity, which may be related to their DNA profile? What is the population from which the criminal may be thought to have come? Could they be another member of the family of the PoI?

Questions such as these and their effect on the interpretation and evaluation of evidence will be discussed in greater detail. First, consider only the evidence of the DNA profile in isolation and one particular locus, *LDLR*. It is no longer used in forensic genetics; it is used here for ease of explanation. Assume the crime was committed in Chicago and that there is eyewitness evidence that the criminal was a Caucasian. Information is available to the investigating officer about the genotypic distribution for the *LDLR* locus in Caucasians in Chicago and is given in Table 1.1. The information about the location of the crime and

Table 1.1 Genotypic relative frequencies for locus *LDLR* amongst Caucasians in Chicago based on a sample of size 200.

Genotype	*AA*	*BB*	*AB*
Relative frequency (%)	18.8	32.1	49.1

Source: From Johnson and Peterson (1999). Reprinted with permissions of ASTM International.

the ethnicity of the criminal is relevant. Genotypic population proportions vary across locations and amongst ethnic groups. A PoI is arrested and a swab is taken for a comparative genetic test. For locus *LDLR* the genotype of the crime stain and that of the PoI correspond. The investigating officer knows a little about probability and works out that the probability of two people chosen at random and unrelated to the suspect having corresponding alleles, using the figures in Table 1.1 as estimates of population proportions, is

$$0.188^2 + 0.321^2 + 0.491^2 = 0.379, \qquad (1.1)$$

(see Section 3.5.1). They are not too sure what this result means. Is it high and is a high value incriminating for the PoI? Is it low and is a low value incriminating? In fact, a low value is more incriminating than a high value.

They think a little more and remembers that, not only do the genotypes correspond, but also that they are both of type *BB*. The proportions of genotypes *AA* and *AB* are not relevant. He works

out the probability that two people chosen at random both have genotype *BB* as

$$0.321^2 = 0.103,$$

(see Section 3.5). He is still not too sure what this means but feels that it is more representative of the information available to him than the previous probability, since it takes account of the actual genotypes of the crime stain and the PoI.

The genotype of the crime stain for locus *LDLR* is *BB*. The genotype of the PoI is also *BB* (if it were not they would not be a PoI). What is the value of this evidence? The discussion earlier suggests various possible answers.

(1) The probability that two people chosen at random have the same genotype for locus *LDLR*. This is 0.379.

(2) The probability that two people chosen at random have the same, pre-specified, genotype. For genotype *BB* this is 0.103.

(3) The probability that one person, chosen at random, has the same genotype as the crime stain. If the crime stain is of group *BB*, this probability is 0.321, from Table 1.1.

The relative merits of these answers will be discussed in Section 3.5 for (1) and (2) and Section 2.4.5 for (3).

The phrase *at random* is taken to include the caveat that the people chosen are unrelated to the PoI. In practice, the phrase *at random* is not

considered in its scientific usage (a person chosen accordingly to a randomising device). Note that, as expressed by Balding and Steele (2015),

> *It is important to keep in mind that in any crime investigation,* random man *is pure fiction: nobody was actually chosen at random in any population, and so probabilities calculated under an assumption of randomly sampled suspects have no direct bearing on evidential weight in actual cases. (p. 160)*

A comment on randomness is also made by Kingston and Kirk (1964) (see pp. 515–516). A discussion about the extension of the concept of 'random man' to the one of 'unrelated person' to the PoI, when dealing with DNA evidence evaluation, is discussed in Milot et al. (2020).

1.3.3 Glass Fragments

Section 1.3.2 discussed an example of the interpretation of the evidence of DNA profiling. Consider now an example concerning glass fragments and the measurement of the refractive index of these.

Example 1.2. As before, consider the investigation of a crime. A window has been broken during the commission of the crime. A PoI is found with fragments of glass on their clothing, similar in refractive index to the broken window. Several fragments are taken for investigation and their refractive index measurements taken.

Note that there is a difference here from Example 1.1, where it was assumed that the crime stain had come from the criminal and been transferred to the crime scene. In Example 1.2 glass is transferred from the crime scene to the criminal. Glass on the PoI need not have come from the scene of the crime; it may have come from elsewhere and by perfectly innocent means. This is an asymmetry associated with this kind of scenario. The evidence is known as *transfer evidence*, as discussed in Section 1.1, because evidence (e.g. blood or glass fragments) has been transferred from the criminal to the scene or *vice versa*. Transfer from the criminal to the scene has to be considered differently from evidence transferred from the scene to the criminal. A full discussion of this is given in Chapters 5 and 6.

Comparison in Example 1.2 has to be made between the two sets of fragments on the basis of their refractive index measurements. The evidential value of the outcome of this comparison has to be assessed. Notice that it is assumed that none of the fragments has any distinctive features and comparison is based only on the refractive index measurements.

Methods for evaluating such evidence were discussed in many papers in the late 1970s and early 1980s Evett (1977, 1978), Evett and Lambert (1982, 1984, 1985), Grove (1981, 1984), Lindley (1977c), Seheult (1978), and Shafer (1982). These methods will be described as appropriate

in Chapters 3 and 7. Knowledge-based computer systems have been developed. See Curran and Hicks (2009) and Curran (2009) for a review of practices in the forensic evaluation of glass and DNA evidence. As an aside, sophisticated systems have been developed to deal with DNA, notably DNA mixtures complexities (i.e. number of donors, peaks heights). Examples are presented and evaluated in Bright et al. (2016), Alladio et al. (2018), and Bleka et al. (2019).

Evett (1977) gave an example of the sort of problem that may be considered and developed a procedure for evaluating the evidence that mimicked the interpretative thinking of the forensic scientist of the time. The case is an imaginary one. Five fragments from a suspect are to be compared with 10 fragments from a window broken at the scene of a crime. The values of the refractive index measurements are given in Table 1.2. The procedure developed by Evett is a two-stage one. It is described here briefly. It is a rather arbitrary and hybrid procedure. While it follows the thinking of the forensic scientist, there are interpretative problems, which are described here, in attempting to provide due value to the evidence. An alternative approach that overcomes these problems is described in Chapter 7.

The first stage is known as the *comparison* stage. The two sets of measurements are compared. The comparison takes the form of the calculation of a statistic, D, say. This statistic provides a

Table 1.2 Refractive index measurements.

Measurements from the window	1.518 44	1.518 48	1.518 44	1.518 50	1.518 40
	1.518 48	1.518 46	1.518 46	1.518 44	1.518 48
Measurements from the PoI	1.518 48	1.518 50	1.518 48	1.518 44	1.518 46

measure of the difference, known as a *standardised* difference, between the two sets of measurements that takes account of the natural variation there is in the refractive index measurements of glass fragments from within the same window. If the absolute value of D is less than (or equal to) some pre-specified value, known as a *threshold* value, then the two sets of fragments are deemed to be similar and the second stage is implemented. If the absolute value of D is greater than the threshold value, then the two sets of fragments are deemed to be dissimilar. The two sets of fragments are then deemed to have come from different sources and the second stage is not implemented. (Note the use here of the word *statistic*, which in this context can be thought of simply as a function of the observations.) A classic example of such an approach is the use of the Student t-test or the modified Welch test for the comparison of means (Welch 1937; Walsh et al. 1996; Curran et al. 2000).

The second stage is known as the *significance* stage. This stage attempts to determine the significance of the finding from the first stage that the two sets of fragments were similar. The significance is determined by calculating the probability of the result that the two sets of fragments were found to be similar, under the assumption that the two sets had come from different sources. If this probability is very low then this assumption is deemed to be

false. The fragments are then assumed to come from the same source, an assumption that places the PoI at the crime scene. This assumption says nothing about how the fragments came to be associated with the PoI. This may have occurred in an innocent manner. See Section 5.3.2 for further discussion of this point in the context of activity level propositions.

The procedure can be criticised on two points. First, in the comparison stage the threshold provides a qualitative step that may provide very different outcomes for two different pairs of observations. One pair of sets of fragments may provide a value of D, which is just below the threshold, whereas the other pair may provide a value of D just above the threshold. The first pair will proceed to the significance stage, the second stage will not. Yet, the two pairs may have measurements, which are close together. The difference in the consequences is greater than the difference in the measurements merits (such an approach is called a *fall-off-the-cliff* effect; see Evett (1991) who attributed this term to Ken Smalldon. Criticisms have been developed in Robertson and Vignaux (1995b). They wrote:

> *This sudden change in decision due to crossing a particular line is likened to falling off a cliff, one moment you are safe, the next dead. In fact, rather than a cliff we have merely a steep slope. Other things being equal, the more similar the samples the stronger the evidence that they had a common origin, and the less similar the samples the stronger the evidence that they came from different sources. (p. 118)*

A related problem is that of cut-off where a decision is taken dependent on whether a statistic is above or below a certain value (see Section 7.7.5).

A better approach, as suggested in the quotation from Robertson and Vignaux (1995b) above and that is described in Section 7.3, provides a measure of the value of the evidence that decreases as the distance between the two sets of measurements increases, subject, as explained later, to the rarity or otherwise of the measurements.

The second criticism is that the result is difficult to interpret. Because of the effect of the comparison stage, the result is not just simply the probability of the evidence, assuming the two sets of fragments came from different sources. A reasonable interpretation, as will be explained in Section 2.4, of the value of the evidence is the effect that it has on the odds in favour of the true guilt of the PoI. In the two-stage approach this effect is difficult to measure. The first stage discards certain sets of measurements, which may have come from the same source and does not discard other sets of measurements which may have come from different sources. The second stage calculates a probability, not of the evidence but of that part of the evidence for which D was not greater than the threshold value, assuming the two sets came from different sources. It is necessary, as is seen later, to compare this probability with the probability of the same result, assuming the two sets came from the same source. There is also an implication in the determination of the

probability in the significance stage that a small probability for the evidence, assuming the two sets came from different sources, means that there is a large probability that the two sets came from the same source. This implication is unfounded; see Section 2.5.1.

A review of the two-stage approach and the development of a Bayesian approach is provided by Curran et al. (2000) and Curran and Hicks (2009).

As with DNA profiling, there are problems associated with the definition of a suitable population from which probability distributions for refractive measurements may be obtained; see, for example, Walsh and Buckleton (1986).

These examples have been introduced to provide a framework within which the evaluation of evidence may be considered. In order to evaluate evidence, something about which there is much uncertainty, it is necessary to establish a suitable terminology and to have some method for the assessment of uncertainty. First, some terminology will be introduced followed by a method for the measurement of uncertainty. This method is *probability*. The role of uncertainty, as represented by probability, in the assessment of the value of scientific evidence will form the basis of the rest of this chapter. A commentary on so-called *knowledge management*, of which this is one part, has been given by Evett (1993b, 2015).

1.4 TERMINOLOGY

It is necessary to have clear definitions of certain terms. First note that the term *evidence* is generally used in the literature and in practice rather than terms such as *finding*, *outcome*, or *material*. The term *evidence* may be confusing. In some legal contexts it can refer to a judicial qualification of a finding. Forensic scientists are interested in the probative value of an observation before it qualifies as *evidence* in a trial. The European Networks of Forensic Science Institutes (ENFSI) Guideline for evaluative reporting (ENFSI 2015) provides definitions of both terms (pp. 19–20). Despite the risk of confusion, the term *evidence* will be used for the material examined by the scientist and for which the value is required.

The crime scene and suspect (associated with a PoI) materials have fundamentally different roles. The assignment of a probability for a correspondence between two randomly chosen sets of materials is not the important issue. One set of materials, crime scene or suspect, can be assumed to have a known source. It is then required to assess the probability of the corresponding material, suspect or crime scene, corresponding in some sense, to the known set of materials, under two competing hypotheses. Examples 1.1 and 1.2 serve to illustrate this.

Example 1.1. (continued) A crime is committed. A bloodstain is found at the scene of the crime.

All innocent explanations for the presence of the stain are eliminated. A PoI is found. Their DNA profile is found to match that of the crime stain. The crime scene material is the crime stain. The suspect material is a swab with saliva. The finding to evaluate is the observed correspondence between the DNA profiles of the crime scene and suspect materials.

Example 1.2 (continued) As before, consider the investigation of a crime. A window has been broken during the commission of the crime. Several fragments are taken for investigation and measurements made of their refractive indices. These fragments, as their origin is known, are sometimes known as *control fragments* and the corresponding measurements are known as *control measurements*. A PoI is found. Fragments of glass are found on their person and measurements of the refractive indices of these fragments are made. These fragments are sometimes known as *recovered fragments*. Their origin is not known. They may have come from the window broken at the crime scene but need not necessarily have done so.

The crime scene material is the fragments of glass and the measurements of refractive index of these at the scene of the crime. The suspect material is the fragments of glass found on the PoI and their refractive index measurements.

Evidence where the source is known will be known as *source* evidence. These fragments of glass will be known as source fragments and the

corresponding measurements will be known as source measurements, as their source is known. An alternative term for this type of evidence was bulk source evidence (Stoney 1991a) but this terminology appears to have fallen into disuse.

A PoI is identified. Fragments of glass are found on their person and measurements of the refractive indices of these fragments are made. Evidence such as this where the evidence has been received and is in particulate form will be known as *transferred particle* evidence. The fragments of glass in this example will be known as *transferred particle* fragments. Their origin is not known. They have been received from somewhere by the PoI. They are particles that have been transferred to the PoI from somewhere. They may have come from the window broken at the crime scene but need not necessarily have done so.

There will also be occasion to refer to the location at which, or the person on which, the evidence was found. Material found at the scene of the crime will be referred to as *crime* evidence. Material found on the suspect's clothing or in the suspect's natural environment, such as their home, will be referred to as *suspect* evidence. Note that this does not mean that the evidence itself is of a suspect nature!

Locard's principle (see Section 1.1) is that every contact leaves a trace. In the earlier examples the contact is that of the criminal with the crime scene. In Example 1.1, the trace is the bloodstain at the crime scene. In Example 1.2, the trace is the fragments of glass that would be removed from the

crime scene by the criminal (and, later, hopefully, be found on their clothing).

The evidence in both examples is *transfer evidence* (see Section 1.1) or sometimes *trace evidence*. Material has been transferred between the criminal and the scene of the crime. In Example 1.1 blood has been transferred *from* the criminal *to* the scene of the crime. In Example 1.2 fragments of glass *may* have been transferred *from* the scene of the crime *to* the criminal. Note that the direction of transfer in these two examples is different. Also, in the first example the blood at the crime scene has been identified as coming from the criminal. Transfer is known to have taken place. In the second example it is not known that glass has been transferred from the scene of the crime to the criminal. The PoI has glass fragments on his clothing but these need not necessarily have come from the scene of the crime. Indeed if the PoI is innocent and has no connection with the crime, the fragments will not have come from the crime scene.

The term *control evidence* has been used to indicate the material whose origin is known. Similarly, the term *recovered* has been used to indicate the material whose origin is unknown.

Alternatively, *questioned* has been used for 'recovered'. See, for example, Brown and Cropp (1987). Also Kind et al. (1979) used the word *crime* for material known to be associated with a crime and *questioned* for material thought to be associated with a crime. All these terms are

ambiguous. The need to distinguish the various objects or persons associated with a crime was pointed out by Stoney (1984).

Definitions given in the particular context of fibre evidence are provided also by Champod and Taroni (1999). The object or person on which traces have been recovered is defined as the *receptor*, and the object or person that could be the source (or one of the sources) of the traces, and which is the origin of the material defined as *known material*, is defined as the *known source*.

Material will be referred to as *control* form where appropriate and to *recovered* or *transferred* particle form where appropriate. In the previous examples, there are two possibilities for the origin of the material which is taken to be known: the scene of the crime and the PoI. One or other is taken to be known, the other to be unknown. The two sets of material are compared. Two probabilities for what is assumed known are determined. One depends on an assumption of common source. The other depends on an assumption of different sources. The two possibilities for the origin of the material that is taken to be known are called *scene-anchored* and *suspect-anchored*, where the word 'anchored' refers to that which is assumed known (Stoney 1991a). The distinction between scene-anchoring and suspect-anchoring is important when determining relevant probabilities (Section 5.3.1.4); it is not so important in the determination of likelihood ratios or Bayes' factors (see Section 2.3.2). Reference to form (source or

receptor) is a reference to one of the two parts of the evidence. Reference to anchoring (scene or suspect) is a reference to a perspective for the evaluation of the evidence.

It is sometimes useful to refer to material found at the scene of a crime as the *crime scene item* and to material found on or about a PoI as the *suspect item*. This terminology reflects the site at which the material was found. It does not indicate the kind of material (bulk or transferred particle form) or the perspective (scene – or suspect – anchored) by which the evidence will be evaluated.

1.5 TYPES OF DATA

A generic name for observations that are made on items of interest, such as bloodstains, refractive indices of glass, etc. is *data*. There are different types of data and some terminology is required to differentiate amongst them. For example, consider shoe types. The observations of interest are the shoe types as observable on surveillance camera recordings and those observable in possession of a PoI. These shoe types are not quantifiable. There is no numerical significance that may be attached to these. The shoe type is a qualitative characteristic. As such, it is an example of so-called *qualitative data*, sometimes known as *categorical data*. The observation of interest is a quality, the shoe type, which has no numerical significance. The different shoe types are sometime known as

categories. The assignation of a shoe to a particular category is called a *classification*. A shoe may be said to be assigned to one of several categories (see the discussion on the definition of *identification* in Section 2.5.9). Other examples of categorical data include types of firearms and makes of cars.

It is not possible to order shoe types and say that one type is larger or smaller than another. However, there are other qualitative data that do have a natural ordering, such as the level of burns on a body. There is not a numerical measure of this but the level of burns may be classified as first, second, third degree, for example. Qualitative data that have no natural ordering are known as *nominal* data. Qualitative data to which a natural ordering may be attached are known as *ordinal* data. An ordinal characteristic is one in which there is an underlying order even though it is not quantifiable. Pain is another such characteristic; level of trauma may be ordered as none, slight, mild, severe, or very severe. The simplest case of nominal data arises when an observation may be classified into one of only two possible categories. For example, consider the magnetism of toner present on printed documents. Some toners are magnetic, whereas others are not. Such data are known as *binary*. Alternatively, the variable of interest, here magnetism, is known as *dichotomous*; it is either present or absent (Biedermann et al. 2016c).

Other types of data are known as *quantitative* data. These may be either counts (known as *discrete* data, since the counts take discrete, integer,

values) or measurements (known as *continuous* data, since the measurements may take any value on a continuous interval).

A violent crime involving several people, victims and offenders, may result in much blood being spilt and many stains from each of several DNA profiles being identified. Then the numbers of stains for each of the different profiles are examples of discrete, quantitative data. Other examples of discrete quantitative data are the number of glass fragments found on a PoI, or the number of gunshot residue particles on hands.

The refractive indices and elemental concentrations of glass fragments are examples of continuous measurements. In practice, variables are rarely truly continuous because of the limits imposed by the sensitivity of the measuring instruments. For example, refractive indices may be measured only to a certain number of decimal places.

Observations, or data, may thus be classified as qualitative or quantitative. Qualitative data may be classified further as nominal or ordinal, and quantitative data may be classified further as discrete or continuous.

1.6 POPULATIONS

'Who is "random man"?' This is the title of a paper by Buckleton et al. (1991). In order to evaluate

evidence, it is necessary to have some idea of the variability or distribution of the evidence under consideration within some population. This population will be called the *relevant* population (and a more formal definition will be given later in Section 6.1.1) because it is the population that is deemed relevant to the evaluation of the evidence. Variability is important because if the PoI did not commit the crime and is, therefore, assumed innocent it is necessary to be able to determine the probability of associating the evidence with them when they are innocent. Surveys of populations are required in order to obtain this information. Surveys for reference data are regularly published in forensic science or legal medicine journals (e.g. *Forensic Science International*, *Science & Justice*, *Journal of Forensic Sciences*, *International Journal of Legal Medicine*); they are widely available to the scientific community.

Care has to be taken when deciding how to choose the relevant population. Buckleton et al. (1991) describe two situations and explain how the relevant population is different for each.

The first situation is one in which there is transfer from the criminal to the crime scene as in Example 1.1 and discussed in greater detail in Section 5.3.1.4. In this situation, the details of any PoI are irrelevant under H_d, the proposition that the PoI was not present at the scene of the crime. Consider a bloodstain at the crime scene that, from the background information I, it is possible to assume is of blood from the criminal.

If the PoI was not present, then clearly some other person must have left the stain. There is no reason to confine attention to any one group of people. In particular, attention should not be confined only to any group (e.g. ethnic group) to which the PoI belongs. However, if there is some information that might cause one to reconsider the choice of population, then that choice may be modified. Such information may come from an eyewitness, for example, who is able to provide information about the offender's ethnicity. This would then be part of the background information I. Further comments on the role of background information I may be found in Section 2.4.4. In general, though, information about DNA profile frequencies would be required from a survey, which is representative of all possible offenders. For evidence of DNA profiles, it is known that age is not a factor affecting a person's profile but that ethnicity is. It is necessary to consider the racial composition of the population of possible offenders (not persons of interest). Normally, it is necessary to study a general population since there will be no information available to restrict the population of possible criminals to any particular ethnic group or groups.

The second situation considered by Buckleton et al. (1991) is possible transfer from the crime scene to the PoI or criminal and discussed further in Section 5.3.2.4. The details of the PoI are now relevant even assuming they were not present at the crime scene. Consider the situation where

there is a deceased victim who has been stabbed numerous times. A PoI, with a history of violence, has been apprehended with a heavy bloodstain on their jacket that is not of their own blood. What is the evidential value in itself, and not considering possible DNA evidence, of the existence of such a heavy bloodstain, not of the blood of the PoI? The probability of such an event (the existence of a heavy bloodstain) if the PoI did not commit the crime needs to be considered.

The PoI may offer an alternative explanation. The jury can then assign a probability to the occurrence of the evidence, given that explanation. The two propositions to be considered would then be

H_p: the blood was transferred during the commission of the crime;

H_d: the explanation of the PoI is true,

and the jury could assess the evidence of the existence of transfer under these two propositions. Evaluation of the evidence of the DNA profile frequencies would be additional to this. The two parts could then be combined using the technique described in Section 5.3.2.

In the absence of an explanation from the PoI, the forensic scientist could conduct a survey of persons as similar as possible to the PoI in whatever are the key features of their behaviour or lifestyle. The survey would be conducted with respect to the PoI since it is of interest to learn

about the transfer of bloodstains for people with their background. In a particular case, it may be that a survey of people of a violent background is needed. One example is that of Briggs (1978) in which 122 suspects who were largely vagrants, alcoholics, and violent sexual deviants were studied. The nature and lifestyle of the PoI determine the type of population to survey. Buckleton et al. (1991) reported also the work of Fong and Inami (1986) in which clothing items from persons of interest, predominantly in offences against the person, were searched exhaustively for fibres that were subsequently grouped and identified.

The idea of a relevant population is a very important one and is discussed further in Section 6.1.1 following the development proposed by Champod et al. (2004). Consider the example of offender profiling, one which is not strictly speaking forensic science but which is still pertinent during an investigation. Consider the application to rape cases. Suppose the profiler is asked to comment on the offender's lifestyle, such as age, marital status, existence and number of previous convictions, and so on, which the profiler may be able to do. However, it is important to know something about the distribution of these in some general population. The question arises here, as in Buckleton et al. (1991) described earlier, as to what is the relevant population. In rape cases, it may not necessarily be the entire male population of the local community. It could be argued that it might be the population of burglars, not so much

because rapists are all burglars first but rather burglars are a larger group of people who commit crimes involving invasion of someone else's living space. Information from control groups is needed, regarding both the distribution of observed traits amongst the general non-offending population and the distribution of similar offences amongst those without the observed traits. Discussions on relevant population have also been published in legal journals, see, for example, Lempert (1991).

1.7 PROBABILITY

1.7.1 Introduction

The interpretation of scientific evidence may be thought of as the assessment of a comparison. The comparison is that between the recovered material (denote this by M_r) and the control material (denote this by M_c). Denote the combination by $M = (M_r, M_c)$. As a first example, consider the bloodstains of Example 1.1. The crime stain is M_r, the recovered evidential material (i.e. evidential material whose source is unknown), and M_c is the genotype of biological material (e.g. blood, saliva swab) taken from the suspect under controlled conditions (i.e. so-called control material whose source is known). From Example 1.2, suppose glass is broken during the commission of a crime. M_c would be the fragments of glass (the control

material) found at the crime scene, M_r would be fragments of glass (the recovered material) found on the clothing of a suspect, and M would be the two sets of fragments.

Qualities, such as genotypes, or measurements, such as the refractive indices of glass fragments, are taken from M. Comparisons are made of the measurements made on recovered and control material. Denote these by E_r and E_c, respectively, and let $E = (E_r, E_c)$ denote the combined set. Comparison of E_r and E_c is to be made and the assessment of this comparison has to be quantified. The totality of the evidence is denoted Ev and is such that $Ev = (M, E)$.

Statistics has developed as a subject in which one of its main concerns is the quantification of the assessments of comparisons. The performance of a new treatment, drug, or fertiliser has to be compared with that of an old treatment, drug, or fertiliser, for example. Two sets of materials, control and recovered, are to be compared. It seems natural that statistics and forensic science should come together, and this has been happening over the last 40 years after strong criticisms from some outstanding quarters. Recall Kirk and Kingston (1964). They remarked that

> *When we claim that criminalistics is a science, we must be embarrassed, for no science is without some mathematical background, however meagre. This lack must be a matter of primary concern to the educator [...]. Most, if not all, of the amateurish efforts of all of us to justify our own evidence interpretations have been deficient in mathematical exactness and philosophical understanding. (pp. 435–436)*

They concluded by affirming that

> *It can be fairly stated that there is no form of evidence whose interpretation is so definite that statistical treatment is not needed or desirable. (p. 437)*

As discussed in Section 1.2, there have been several books describing the role of statistics in the law. Until the first edition of this book, there had been none concerned with statistics and the evaluation of scientific evidence. Two factors may have been responsible for this.

First, there was a lack of suitable data from relevant populations. There was a consequential lack of a baseline against which measures of typicality of any characteristics of interest may be determined. One exception are the reference data that have been available for many years on allele frequencies for DNA analysis amongst certain populations. Not only has it been possible to say that the DNA of a PoI corresponded[4] to that of a stain found at the scene of a crime, but also that this profile is only present in, say, 0.01% of the population. Now these have been superceded by results of surveys of allele frequencies in various populations. Announcements of population data are published regularly in peer-reviewed journals such as *Forensic Science International: Genetics* and the *International Journal of Legal Medicine.*

[4]The term *correspond* is used here rather than the more commonly used *match*. The term 'match' suggests certainty or identity which is not always the case with trace evidence. Use of the term 'correspond' emphasises this and is a reminder to be careful with interpretation. See Section 2.5.11 and Friedman (1996) for further comment.

Also, data collections exist for the refractive index of glass fragments found at random on clothing and for transfer and persistence parameters linked to glass evidence; see, for example, Curran et al. (2000), O'Sullivan et al. (2011), and Jackson et al. (2013). Contributions towards characterising the rarity of different fibre types have also been published since the late 1990s; for a review, see Palmer (2016).

Secondly, the approach adopted by forensic scientists in the assessment of their evidence has been difficult to model. The approach has been one of comparison and significance. Characteristics of the control and recovered items are compared. If the examining scientists believe them to be similar, the typicality, and hence the significance of the similarity, of the characteristics is then assessed. This approach is what has been modelled by the two-stage approach of Evett (1977), described briefly in Section 1.3.3 and in fuller detail in Chapter 3. However, interpretation of the results provided by this approach is difficult.

Then, in a classic paper, Lindley (1977c) described an approach that was easy to justify, to implement, and to interpret. It combined the two parts of the two-stage approach into one statistic and is discussed in detail in Section 7.4.3. The approach compares two probabilities, the probability of the evidence, assuming one proposition to be true (e.g. that a PoI is the source of the evidence), and the probability of the evidence, assuming another, mutually exclusive,

proposition to be true (e.g. that the PoI is not the source of the evidence). Note that some people use the term *hypothesis* rather than *proposition*; the authors will endeavour to use the term *proposition* as they believe this reduces the risk of confusion of their ideas with the ideas of *hypothesis testing* associated with the alternative term. A proposition is interpreted here as an assertion or statement that, for example, a particular outcome has occurred or a particular state of nature occurs.

This approach implies that it is not enough for a prosecutor to show that there is a low probability to observe the evidence if a PoI is innocent. It should also be more probable to observe the evidence if the PoI is truly guilty. Such an approach has a good historical pedigree (Good, 1950, and also Good, 1991, for a review) yet it had received very little attention in the forensic science literature, even though it was clearly proposed at the beginning of the twentieth century (Taroni et al. 1998; Champod et al. 1999), and earlier by Peirce (1878). It is also capable of extension beyond the particular type of example discussed by Lindley, as will be seen by the discussion throughout this book, for example, in Chapters 6 and 7.

However, in order to proceed it is necessary to have some idea about how uncertainty can be measured. This is best done through probability (Lindley 1991, 1985, 2014). This central role for probability in evidence evaluation is supported by the ENFSI. In the ENFSI Guideline for evaluative

reporting in forensic science,[5] is reported, at page 6 (under point 2.3), that:

> *Evaluation of forensic science findings in court uses probability as a measure of uncertainty. This is based upon the findings, associated data and expert knowledge, case specific propositions and conditioning information.*

where the term 'findings' denotes 'evidence' in our usage.

1.7.2 A Standard for Uncertainty

An excellent description of probability and its role in forensic science has been given by Lindley (1991). Lindley's description starts with the idea of a standard for uncertainty. He provides an analogy using the concept of balls in an urn. Initially, the balls are of two different colours, black and white. In all other respects, size, weight, texture, etc., they are identical. In particular, if one were to pick a ball from the urn, without looking at its colour, it would not be possible to tell what colour it was. The two colours of balls are in the urn in proportions b and w for black and white balls, respectively, such that $b + w = 1$. For example, if there were 10 balls in the urn of which 6 were black and 4 were white, then $b = 0.6, w = 0.4$, and $b + w = 0.6 + 0.4 = 1$.

[5]The 2015 Guidelines for evaluative reporting can be found at http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

The urn is shaken up and the balls thoroughly mixed. A ball is then drawn from the urn. Because of the shaking and mixing, it is assumed that each ball, regardless of colour, is equally likely to be selected. Such a selection process, in which each ball is equally likely to be selected, is known as a random selection, and the chosen ball is said to have been chosen *at random*.

The ball, chosen at random, can be either black, an event that will be denoted B, or white, an event that will be denoted W. There are no other possibilities; one and only one of these two events has to occur. The uncertainty of the event B, the drawing of a black ball, is related to the proportion b of black balls in the urn. If b is small (close to zero), B is unlikely. If b is large (close to 1), B is likely. A proportion b close to $1/2$ implies that B and W are about equally likely. The proportion b is referred to as the probability of obtaining a black ball on a single random drawing from the urn. In a similar way, the proportion w is referred to as the probability of obtaining a white ball on a single random drawing from the urn.

Notice that on this simple model probability is represented by a proportion. As such it can vary between 0 and 1. A value of $b = 0$ occurs if there are no black balls in the urn, and it is, therefore, impossible to draw a black ball from the urn. The probability of obtaining a black ball on a single random drawing from the urn is zero. A value of

$b = 1$ occurs if all the balls in the urn are black. It is certain that a ball drawn at random from the urn will be black. The probability of obtaining a black ball on a single random drawing from the urn is one. All values between these extremes of 0 and 1 are possible (by considering very large urns containing very large numbers of balls).

A ball has been drawn at random from the urn. What is the probability that the selected ball is black? The event B is the selection of a black ball. Each ball has an equal chance of being selected. The colours black and white of the balls are in the proportions b and w, respectively. The proportion, b, of black balls corresponds to the probability that a ball, drawn in the manner described (i.e. at random) from the urn is black. It is then said that the probability a black ball is drawn from the urn, when selection is made at random, is b. Some notation is needed to denote the probability of an event. The probability of B, the drawing of a black ball, is denoted $\Pr(B)$ and similarly $\Pr(W)$ denotes the probability of the drawing of a white ball. Then it can be written that $\Pr(B) = b$ and $\Pr(W) = w$. Note that

$$\Pr(B) + \Pr(W) = b + w = 1.$$

This concept of balls in an urn can be used as a reference for considering uncertain events. The methodology has been described as follows (Lindley 2006):

> *Your*[6] probability *of the uncertain event of rain tomorrow is the fraction of [black] balls in an urn from which the withdrawal of a [black] ball at random is an event of the same uncertainty for you as that of the event of rain. [. . .] You are invited to compare that event with the standard, adjusting the number of [black] balls in the urn until you have the same beliefs in the event and in the standard. Your probability for the event is then the resulting fraction of [black] balls. (p. 35)*

Another example concerns a hypothetical sporting event. Let R denote the uncertain event that the England football team will win the next major international football championship. Let B denote the uncertain event that a black ball will be drawn from the urn. A choice has to be made between R and B, and this choice has to be ethically neutral. If B is chosen and a black ball is drawn from the urn then a prize is won. If R is chosen and England do win the championship the same prize is won. The proportion b of black balls in the urn is known in advance. Obviously, if $b = 0$ then R is the better choice, assuming, of course, that England do have some non-zero probability of winning the championship. If $b = 1$ then B is the better choice. Somewhere in the interval $[0, 1]$, there is a value of b, b_0 say, where the choice does not matter to You. You are indifferent as to whether R or B is chosen. If B

[6]Note the use of the capitalised words 'You' and 'Your' in this quotation. This is a rhetorical device to help readers keep in mind that probabilities are their own degrees of belief based on the information they have at the time they make the judgement.

is chosen $\Pr(B) = b_0$. Then it said that $\Pr(R) = b_0$ also. In this way, the uncertainty in relation to any event can be measured by a probability b_0, where b_0 is the proportion of black balls, which leads to indifference between the two choices, namely, the choice of drawing a black ball from the urn and the choice of the uncertain event in whose probability one is interested.

Notice, though, that there is a difference between these two probabilities. By counting, the proportion of black balls in the urn can be determined precisely. Probabilities of other events such as the outcome of the toss of a coin or the roll of a die are also relatively straightforward to determine, based on assumed physical characteristics such as fair coins and fair dice. Let H denote the event that when a coin is tossed it lands head uppermost. Then, for a fair coin, in which the outcomes of a head H or a tail T at any one toss are considered as equally likely, the probability the coin comes down head uppermost is $1/2$. Let F denote the event that when a die is rolled it lands 4 uppermost. Then, for a fair die, in which the outcomes $1, 2, \ldots, 6$ at any one roll are equally likely, the probability the die lands 4 uppermost is $1/6$.

Probabilities relating to the outcomes of sporting events, such as football matches or championships or horse races, or to the outcome of a civil or criminal trial, are rather different in nature. It may be difficult to decide on a particular value for b_0. The value may change as evidence accumulates such

as the results of particular matches and the fitness or otherwise of particular players, or the fitness of horses, the identity of the jockey, the going of the race track, etc. Also, different people may attach different values to the probability of a particular event.

These kinds of probability – as briefly specified before in Section 1.3 – are sometimes known as *subjective* or *personal* probabilities; see de Finetti (1933), Savage (1954), Good (1959), DeGroot (1970), and the more recent publication by Kadane (2011). Another term is *measure of belief* since the probability may be thought to provide a measure of one's belief in a particular event. A philosophical discussion on the use of those terms is given in Lucena-Molina (2016, 2017). Despite these difficulties the arguments concerning probability still hold. Given an event R whose outcome is uncertain, the probability that R occurs, $\Pr(R)$, is defined as the proportion of black balls b_0 in the urn such that if one had to choose the outcome B (the event that a black ball was chosen) where $\Pr(B) = b_0$ and the outcome R then one would be indifferent to which one was chosen. There are difficulties but the point of importance is that a standard for probability exists. An extended comment on subjective probabilities is given in Sections 1.7.5–1.7.7.

A use of probability as a measure of belief is described in Section 1.7.5 where it is used to represent *relevance*. The differences and similarities in the two kinds of probability discussed earlier and

their ability to be combined have been referred to as a duality (Hacking 1975).

It is helpful also to consider two quotes concerning the relationship amongst probability, logic and consistency, both from Ramsey (1931).

> *We find, therefore, that a precise account of the nature of partial beliefs reveals that the laws of probability are laws of consistency, an extension to partial beliefs of formal logic, the logic of consistency. They do not depend for their meaning on any degree of belief in a proposition being uniquely determined as the rational one; they merely distinguish those sets of beliefs which obey them as consistent ones. (p. 182)*

> *We do not regard it as belonging to formal logic to say what should be a man's expectation of drawing a white or black ball from an urn; his original expectations may within the limits of consistency be any he likes; all we have to point out is that if he has certain expectations he is bound in consistency to have certain others. This is simply bringing probability into line with ordinary formal logic, which does not criticise premises but merely declares that certain conclusions are the only ones consistent with them. (p. 189)*

In brief, a person is entitled to their own measures of belief, but must be consistent with them. Ramsey's remarks relate to the appropriateness of a set of probabilities held by a particular individual. This appropriateness needs to be checked. Probability values need to be expressed in an operational way that will also make clear what coherence means and what coherent conditions are. De Finetti (1976) framed the operational perspective as follows:

> *However, it must be stated explicitly how these subjective probabilities are defined, i.e. in order to give an operative (and not an empty verbalistic) definition, it is necessary to indicate a procedure, albeit idealised but not distorted, an (effective or conceptual) experiment for its measurement. (p. 212)*

Therefore, one should keep in mind the distinction between the *definition* and the *assessment* of probability. A description of de Finetti's perspective has been published by Dawid and Galavotti (2009).

One way in which these expressions can be checked is to measure probabilities maintained by an individual in terms of bets the individual is willing to accept. An alternative to consideration of balls in an urn is to consider two lotteries. An individual probability can be determined using a process known as *elicitation*. In this context, elicitation is the comparison of two lotteries of the same price. Consider a situation in which it is of interest to determine a probability for rain tomorrow. This example can be found in Winkler (1996). There are two lotteries:

- Lottery A: Win £100 with probability 0.5 or win nothing with probability 0.5.
- Lottery B: Win £100 if it rains tomorrow or win nothing if it does not rain tomorrow.

In this situation, it is reasonable to assume a person would choose that lottery which, in their opinion, presents the greater probability of winning the prize. If lottery B is preferred, then this indicates that one considers the probability of

rain tomorrow to be greater than 0.5. Similarly, a choice of lottery A implies the probability of rain tomorrow is less than 0.5. Additionally, in a case in which one is indifferent between the two lotteries, one's probability for rain tomorrow equates with the probability of winning the prize in lottery A. Therefore, a procedure can be devised in which the probability of winning lottery A is adjusted so that the individual, whose probability for a proposition of interest is to be elicited, is indifferent with respect to lotteries A and B. In a similar manner, the personal probability of an individual for any event of interest can be elicited.

The possibility that subjective degrees of belief may be represented in terms of betting rates in lotteries or in the relative frequency of balls in an urn is often put forward as support for an argument that requires subjective degrees of belief to satisfy the laws of probability. This requirement is satisfied with the notion of *coherence* that has the normative role of forcing people to be honest and to make the best assessment of their own measure of belief.

De Finetti (1931a) showed that coherence, a simple economic behavioural criterion, implies that a given individual should avoid a combination of probability assignments that is guaranteed to lead to loss. All that is needed to ensure such avoidance is for uncertainty to be represented and manipulated using the theory of probability. In this context, the possibility of representing subjective degrees of belief in terms of betting odds is often

forwarded as part of a line of argument to require that subjective degrees of belief should satisfy the laws of probability. This line of argument takes two parts. The first is that betting odds should be coherent, in the sense that they should not be open to a sure-loss contract. The second part is that a set of betting odds is coherent if and only if it satisfies the laws of probability. The Dutch Book argument encompasses both parts: the proof that betting odds are not open to a sure loss contract if and only if they are probabilities is called the *Dutch book theorem*. Thus, if an individual translates their state of knowledge in such a manner that the assigned probabilities, as a whole, do not respect the laws of probability (standard probability axioms), then their assignments are not coherent. In this context, such incoherence is also called *logical imprudence*. An example can be found in Section 1.7.6.

1.7.3 Events

The outcome of the drawing of a ball from the urn was called an event. If the ball was black, the event was denoted B. If the ball was white, the event was denoted W. It was not certain which of the two events would happen: would the ball be black, event B, or white, event W? The degree of uncertainty of the event (B or W) was measured by the proportion of balls of the appropriate colour (B or W) in the urn and this proportion was called the probability of the event (B or W). In general,

for an event R, Pr(R) denotes the probability that R occurs.

As underlined by Lindley (2014), events may have happened (past events), may be relevant at the present time (present events), or may happen in the future (future events).

> *There are some things that you [. . .] know to be true, and others that you know to be false; yet, despite this extensive knowledge that you have, there remain many things whose truth or falsity is not known to you. We say that you are* uncertain *about them. You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. (p. xi)*

So, a person may be uncertain about each of these three types of events. Such uncertainty can be expressed by probability.

- *Past event*: A crime is committed and a bloodstain with a particular DNA profile is found at the crime scene. A PoI is found. The event of interest is that the suspect is the source of the stain at the crime scene. Though the PoI either is or is not the source of the stain the knowledge of it is incomplete and hence there is uncertainty about this event. This uncertainty can be expressed by probability.
- *Present event*: A PoI is identified. The event of interest is that they have a particular DNA profile (e.g. Y-STR haplotype). Again, before the result of a DNA analysis is available, this knowledge is incomplete.

- *Future event*: The event of interest is that it will rain tomorrow.

All of these events are uncertain and have probabilities associated with them. Notice, in particular, that even if an event has happened, its actual outcome may be unknown so that knowledge about it is incomplete. The probability the PoI is the source of the stain at the crime scene requires consideration of many factors, including the possible location of the PoI at the crime scene and the properties of transfer of blood from a person to a site. With reference to the gene expression, consideration has to be given to the proportion of people in some population with that gene expression. Probabilistic statements are common with weather forecasting. Thus, it may be said, for example, that the probability it will rain tomorrow is 0.8 (though it may not always be obvious what this means).

1.7.4 Classical and Frequentist Definitions of Probability and Their Limitations

The *classical* definition of probability defines it as the ratio of the number of favourable cases to the total number of possible cases, provided that all cases are equally probable. There is an obvious circularity to this definition. The statement does not define probability, it only offers a way by which it may be evaluated.

The *frequentist* definition of probability is the limit of the relative frequency of a target event that has occurred in a large number of trials, as the number of trials increases to infinity, with the important and unrealistic assumption that the trials are repeated under identical conditions. This definition limits the range of applications since if frequency is to be used as a measure of probability, it must be possible to repeat the underlying experiment a large number of times under identical conditions. Consider a scenario in which a coin is tossed. This definition of probability is equivalent to the assessment of the probability of a head (or tail) by imagining that the act of tossing a coin is able to be repeated a large number of times under identical conditions (e.g. with the same force). The number of heads is observed and the ratio of heads to the total number of tosses is taken as an estimate of the probability of a head for that coin. The frequentist definition of probability is inconceivable operationally for applications in forensic science. A well-known challenge to the frequentist view in the context of criminal law is given by Lindley (1991).

> *There is nothing wrong with the frequency interpretation or chance. It has not been used in this treatment because it is often useless. What is the chance that the defendant is guilty? Are we to imagine a sequence of trials in which the judgements, 'guilty' or 'not guilty', are made and the frequency of the former found? It will not work because it confuses the judgement of guilt, but, more importantly, because it is impossible to conceive of a suitable sequence.*

> *Do we repeat the same trial with a different jury; or with the same jury bur different lawyers; or do we take all Scottish trials; or only Scottish trials for the same offence? The whole idea of chance is preposterous in this context. (p. 48)*

The example makes clear that a definition of probability based on the long-run relative frequency of an event is inapplicable in many situations arising in real life. There are implicit assumptions that must apply in each of the classical and frequentist definitions. These assumptions are, first, that according to our state of knowledge, all cases are equally likely, and, second, that it is theoretically possible to perform an experiment a large number of times under identical conditions. Use of these assumptions to assign a numerical value to a probability implies a judgement that these assumptions are satisfied. A definition of probability that seeks to avoid subjectivity is based on an acceptance of assumptions that are inherently subjective. The frequentist view presumes the possibility of the performance of a long sequence of experiments under identical conditions, with each experiment being physically independent of all other experiments. These assumptions are typically unachievable in many different applied contexts such as history, law, economics, medicine, and, especially, forensic science. In these contexts, the events of interest are usually not the result of repetitive or replicable processes. On the contrary, they are unique.

This aspect has been explicitly underlined by Kingston and Kirk (1964) in the area of forensic science. The authors wrote:

> *There are many philosophically oriented fundamental ideas of probability. Perhaps the most practical basic approach to the subject lies in the concept of frequency, in which statements of probability express the relative frequencies of repeated events. [. . .] For practical use in criminalistics, it is of little interest what might happen in a long series of trials; the crime is committed only once. Of what use, then, is the above frequency concept? The answer to this lies in another way of looking at probability, which is to consider it as a degree of belief. (p. 514)*

Such complications do not arise with the subjective interpretation of probability because that interpretation does not consider probability as a feature of the external world. Instead probability is understood as a notion that describes the relationship between a person (e.g. You, the reader) who makes a statement of uncertainty and the real world to which that statement relates and in which the person acts. With the subjective concept of probability, it is therefore very reasonable to assign a probability to events that are not repeatable, for example, as in a given judicial context. An extended discussion on the limitations of the classical and frequentist definitions of probability can be found in Taroni et al. (2018b).

1.7.5 Subjective Definition of Probability

In forensic science, it is often emphasised that there is a real paucity of numerical data, so that the

numerical evaluation of evidence (Section 2.3.1) is sometimes very difficult. Examples of this difficulty are the numerical assessments of parameters such as transfer or persistence probabilities (see Sections 6.2.3 and 6.2.4) or even the relevance of a piece of evidence (see Section 6.3). The Bayesian approach considers probabilities as measures of belief (also called subjective probabilities) since such probabilities may be thought of as measures of one's belief in the occurrence of a particular event. The approach allows scientists to assign their probabilities, not only by certified knowledge and experience, but also by any data relevant for the event of interest, such as knowledge of an event that is often available in terms of a relative frequency. This specific relationship between relative frequency and probability is discussed in Section 1.7.6. Note that *frequency* is a term that relates to data and *probability* is a term that relates to personal belief.

Jurists are also interested in probabilistic reasoning using subjective probabilities, notably probabilities related to the credibility of witnesses and the conclusions that might be drawn from their testimony.

Any kind of uncertainty is assessed in the light of the knowledge possessed at the time of the assessment. This idea is not new. The Italian mathematician Bruno de Finetti (de Finetti, 1931a) defined probability – the measure of uncertainty – as a degree of belief, insisting that probability is conditional on the status of information of the subject who assesses it. So, if a given person, *S*, say, is interested in the probability of

an event, E, say, that person's probability $\Pr(E)$ should be written as $\Pr(E \mid I_{s,t})$ where $I_{s,t}$ is the information available to person S at time t.

Subjective probability may be found in many scientific areas (Press and Tanur 2001). In physics, for example, Schrödinger (1947) wrote

> *Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence and thus different numerical probabilities may be attached to the same event. (p. 53)*

He then added that

> *Thus, whenever we speak loosely of the probability of an event, it is always to be understood: probability with regard to a certain given state of knowledge. (p. 54)*

The same perspective is expressed by de Finetti in his famous sentence 'Probability does not exist' (in things) (de Finetti 1975, p. x). Probability is not something that can be known or not known, probabilities are states of mind, not states of nature. This aphorism can also be found in previous statistical and philosophical literature (i.e. de Morgan 1838; Jaynes 2003; Jevons 1913; Maxwell 1990). For example, Jevons (1913) wrote

> *Probability belongs wholly to the mind. This is proved by the fact that different minds may regard the very same event at the same time with widely different degrees of probability [. . .] Probability thus belongs to our mental condition, to the light in which we regard events, the occurrence or non-occurrence of which is certain in themselves. (p. 198)*

Probability is a fact about one's state of mind, not a fact about a phenomenon.

In summary, a person's assessment of their degree of belief (subjective probability) in the truth of a given statement or in the occurrence of an event (i) depends on information, (ii) may change as the information changes, and (iii) may differ from the assessment of others because different individuals may have different information or assessment criteria. In Savage's words (Savage 1954)

> *Probabilistic views hold that probability measures the confidence that a particular individual has in the truth of a particular proposition, for example, the probability that it will rain tomorrow. These views postulate that the individual concerned is in some way 'reasonable', but they do not deny the possibility that two reasonable individuals faced with the same information may have different degrees of confidence in the truth of the same proposition. (p. 3)*

The only constraint in the assessment – as noted in Section 1.7.2 – is that it must be coherent. Coherence may be understood through consideration of subjective probability in terms of betting, for example, on the outcome of a horse race. For the probabilities on winning for each horse in a race to be coherent, the sum of the probabilities over all the horses must be 1. This property characterises a 'reasonable individual'. An example is presented in Section 1.7.6.

For a historical and philosophical discussion of subjective probabilities and a commentary on the work of de Finetti and Savage in the middle

of the twentieth century, see Lindley (1980), Lad (1996), Taroni et al. (2001), Dawid (2004), Dawid and Galavotti (2009), Galavotti (2016, 2017), and Zynda (2016).

Savage, like de Finetti, viewed a personal probability as a numerical measure of the confidence a person has in the truth of a particular proposition. This opinion is viewed with scepticism today and was viewed with scepticism then (Savage 1967), as illustrated by Savage (1954).

> *I personally consider it more probable that a Republican president will be elected in 1996 than it will snow in Chicago sometime in the month of May, 1994. But even this late spring snow seems to me more probable than that Adolf Hitler is still alive. Many, after careful consideration, are convinced that such statements about probability to a person mean precisely nothing or, at any rate, that they mean nothing precisely. At the opposite extreme, others hold the meaning to be so self-evident [. . .]. (p. 27)*[7]

1.7.6 The Quantification of Probability Through a Betting Scheme

The introduction of subjective probability through a betting scheme is straightforward. The concept is based on hypothetical bets (Scozzafava 1987):

> *The force of the argument does not depend on whether or not one actually intends to bet, yet a method of evaluating probabilities making one a sure loser if he had to gamble (whether or not he really will act so) would be suspicious and unreliable for any purposes whatsoever. (p. 685)*

[7]Part of this sentence is also reported by Kadane and Schum (1996, p. 160).

Consider a proposition E that can only take one of two values, namely, 'true' and 'false'. There is a lack of information on the actual value of E and an operational system is needed for the quantification of the uncertainty about E imparted by the lack of information. A value $p = \Pr(E)$ is regarded as an amount to be paid to bet on E with the conditions that a unit amount will be paid if E is true and nothing will be paid if E is false. In other words, p is the amount to be paid to obtain an amount equal to the value of E, that is associating the value 1 with 'true' and the value 0 with 'false'. This idea was expressed by de Finetti (1940) in the following terms.

> *The probability of event E is, according to Mr NN, equal to 0.37, meaning that if the person was forced to accept bets for and against event E, on the basis of the betting ratio p which he can choose as he pleases, this person would choose* $p = 0.37$. *(p. 113)*[8]

Coherence, as briefly described in Section 1.7.2, is defined by the requirement that the choice of p does not make the player a certain loser or a certain winner. Denote an event which is certain, sometimes known as a *universal set*, as Ω and an event which is impossible, sometimes known as the *empty set*, as ϕ so that if $E \neq \Omega$ and $E \neq \phi$ the two possible gains are

$$G_1 = (-p + 1) \text{ if } E \text{ occurs;}$$

$$G_2 = -p \text{ if } E \text{ does not occur.}$$

[8]English version of the paper reprinted in Monari and Cocchi (1993).

When $E = \Omega$ or $E = \phi$, there is no uncertainty in the outcome of the corresponding bet and so the coherence (in the absence of uncertainty) requires the respective gains to be zero. The values of the gains are therefore

$$G(\Omega) = -p + 1 = 0 \text{ and } G(\phi) = -p = 0.$$

This happens when $p = 1$ for $E = \Omega$ and $p = 0$ for $E = \phi$. Therefore if the subjective probability of E, that represents our degree of belief on E, is defined as an amount $p = \Pr(E)$, which makes a personal bet on the event or proposition E coherent, then the probability $\Pr(E)$ satisfies two conditions.

(1) $0 \leq \Pr(E) \leq 1$;

(2) $\Pr(\phi) = 0, \Pr(\Omega) = 1$.

Consider the case of n possible bets on events $E_1, \ldots, E_n$ that partition Ω; i.e. $E_1, \ldots, E_n$ are mutually exclusive and exhaustive (Scozzafava 1987, p. 686). Let $\Pr(E_i), i = 1, \ldots, n$, be the amount paid for a coherent bet on E_i. These bets can be regarded as a single bet on Ω with amount $\Pr(E_1) + \cdots + \Pr(E_n)$. Another condition may be specified from the requirement of coherence, namely

(3) $\Pr(E_1) + \cdots + \Pr(E_n) = 1$.

These conditions are the axioms of probability. Further details are given by de Finetti (1931b) and in

Section 1.7.8. An example of a *Dutch book* is given to examine if a given person assigns subjective probabilities coherently. Consider a horse race with three horses, A, B, and C. A bookmaker offers probabilities of $1/4$, $1/3$, and $1/2$, respectively, for these horses to win. Note that these probabilities add up to more than 1 and so violate condition 3. The corresponding odds are 3 to 1 against winning, 2 to 1 against winning and 'evens'.

The relationship between odds and probability is described briefly here with fuller details given in Chapter 2. An event with probability p of occurring has odds O of happening where $O = p/(1-p)$. Conversely, an event that has odds of O to 1 of happening has a probability of $O/(O+1)$ of happening and an event that has odds of O to 1 of not happening has a probability of $1/(O+1)$ of happening. Odds of 'evens' correspond to $O = 1$ or $p = 1/2$.

Suppose the odds offered by the bookmaker are accepted by the person. Thus, their beliefs do not satisfy the additivity law of probability (condition 3). If any single bet is acceptable, they can all be accepted. This is equivalent to a bet on the certain event that one of A, B, or C wins the race. The individual should therefore expect to break even on the outcome of the race; their winnings will equal their initial stake. Of course, it does not make sense to bet on the certain event as there should then be nothing to win or lose. This assumes the odds

are fixed to satisfy condition 3. However, in this example, the odds do not satisfy condition 3 and the person will not break even. Suppose the following bets are placed: £3000 on *A* to win, £4000 on *B* to win, and £6000 on *C* to win; i.e. £13 000 in total. If *A* wins the bookmaker pays out £12 000, the original £3000 bet and another £9000 in accordance with the odds of 3 to 1 against. If *B* wins, the bookmaker also pays out £12 000, the original £4000 bet and another £8000 in accordance with the odds of 2 to 1 against. If *C* wins the bookmaker again pays out £12 000, the original £6000 bet and another £6000 in accordance with the odds of evens. Regardless of which horse has won the race, the individual has paid out £13 000 and receives £12 000 in winnings, thus incurring a loss of £1000. This situation is known as a Dutch book. The odds quoted did not satisfy condition 3. Conversely, if the set of odds determine probabilities that add up to less than 1, then the bookmaker will lose money. It would be incoherent for such odds to be set.

Judgements are required in all aspect of scientific investigation. The elicitation of probability distributions for uncertain quantities represents a challenging work for scientists and decision-makers. O'Hagan (2019) recently wrote:

> *Subjective expert judgments play a part in all areas of scientific activity, and should be made with the care, rigour, and honesty that science demands. (p. 80)*

A discussion can be found in Section 1.7.7.

1.7.7 Probabilities and Frequencies: The Role of Exchangeability

It is not uncommon for subjective (or personal) probabilities to be considered as a synonym for arbitrariness. This is not so; the use of subjectivism does not mean the use of acquired knowledge that is often available for consideration of relative frequencies is neglected. The main source of misunderstanding is concerned with the relationship between *frequencies* and *beliefs*. The two terms are, unfortunately, often regarded as equivalent since frequency data can be used to inform probabilities (Lindley 1991) but they are not equivalent. Dawid and Galavotti (2009, p. 100) quoted de Finetti's view:

> *every probability evaluation essentially depends on two components: (1) the objective component, consisting of the evidence of known data and facts; and (2) the subjective component, consisting of the opinion concerning unknown facts based on known evidence.*

As emphasised more recently by D'Agostini (2016)

> *It is a matter of fact that relative frequency and probability are somehow connected within probability theory, without the need for identifying the two concepts. (p. 13)*

It is reasonable to use relative frequencies to inform measures of belief and the relationship takes the form of a mathematical theorem, de Finetti's *Representation theorem*. According to

the theorem, the convergence of one's personal probability towards the value of observed frequencies, as the number of observations increases, is a logical consequence of Bayes' theorem if a condition called *exchangeability* is satisfied by the degrees of belief prior to the observations (Dawid 2004).

As an illustration of the connection between frequency and probability, consider again an urn containing a certain number of balls, indistinguishable except by their colour, which is either white or black, and the number of balls of each colour being known. The extraction of a ball from this urn defines an experiment having two and only two possible outcomes that are generally denoted as success (say, the withdrawal of a white ball) or failure (say, the withdrawal of a black ball). Let W denote the event 'a white ball is extracted'. Under the circumstances that balls are all indistinguishable from each other except for the colour, the subjective probability to extract a white ball can be assessed as the known *proportion* θ of white balls, that is, $\Pr(W \mid \theta) = \theta$. Assuming the urn contains a large number of balls, so that the extraction of a few balls does not alter its composition substantially, individual draws (i.e. sampling[9]) will be considered as with

[9]Note that the use of the term 'sample' in this context is one of a purely technical nature in statistics and has nothing to do with the widespread but inappropriate use of the same term for designating physical trace material recovered or collected in a forensic science context. In particular, the seizure (e.g. at crime scenes) and

replacement and the probability of extracting a white ball at subsequent withdrawals will still be θ, independently on previous observations. In this way one realises a series of so-called *Bernoulli trials* (Section A.2.1), where the outcome of each trial has a constant probability independent from previous outcomes.

Suppose now the observer does not know the absolute value of balls present, nor the proportion that are of each colour. De Finetti (1931a) showed that every series of experiments having two and only two possible outcomes that can be taken as exchangeable (i.e. the probability assigned to the outcomes of a sequence of trials is invariant to permutation) can be represented as random withdrawals from an urn of unknown composition. If one can assess one's uncertainty in such a way that labelling of the trials is not relevant, then it can be proved that as the number of observations increases the relative frequencies of successes (i.e. the relative frequency of white balls) tend to a limiting value that is the proportion θ of white balls. A *subjective* assessment about the outcome of a sequence of Bernoulli trials is equivalent to placing a prior distribution on θ. According to this, one only needs to model a prior distribution $\Pr(\theta)$ for the possible values that θ might take: personal beliefs concerning the colour of the next

analysis of trace material has to deal with the material as it is, irrespective of its condition; there is no such thing as randomisation, for example.

ball extracted can be computed as

$$\begin{aligned}\Pr(W) &= \int_{\theta} \Pr(W \mid \theta) \Pr(\theta)\, d\theta \\ &= \int_{\theta} \theta \Pr(\theta)\, d\theta. \end{aligned} \tag{1.2}$$

The introduction of a prior probability distribution modelling personal belief about θ may seem, at first sight, in contradiction with statements that probability is a single number. One can have probabilities for events, or probabilities for propositions, but not probabilities of probabilities, otherwise one would have an infinite regression (de Finetti 1976). Confusion may arise from the fact that parameter θ is generally termed as *'probability of success'*. However, it is worth noting that, although it is effectively a probability, it represents a *chance* rather than a *belief*.

A set of observations $x_1, \ldots, x_n$ is said to be exchangeable – for you, given a knowledge base – if their joint distribution is invariant under permutation. A formal definition is as follows (Bernardo and Smith 2000):

> *The random quantities* $x_1, \ldots, x_n$, *are said to be judged exchangeable under a probability measure* Pr *if the implied joint degree of belief distribution satisfies* $\Pr(x_1, \ldots, x_n) = \Pr(x_{\pi(1)}, \ldots, x_{\pi(n)})$ *for all permutations* π *defined on the set* $\{1, \ldots, n\}$. *(p. 169)*

Practically, consider the following hypothetical case example. A laboratory receives a consignment of discrete items whose attributes may be relevant

within the context of a criminal investigation. The laboratory is requested to conduct analyses in order to gather information that should allow an inference to be drawn, for example about the proportion of items in the consignment that are of a certain kind (e.g. counterfeit products). The term 'positive' is used here to refer to the presence of an item's property that is of interest (e.g. counterfeit); otherwise the result of the analysis is termed 'negative'. This allows the introduction of a random variable X that takes the value 1 (i.e. success) if the analysed unit is positive and 0 (i.e. failure) otherwise. This is a generic type of case that applies well to many situations, such as surveys or, more generally, sampling procedures conducted to infer the proportion of individuals or items in a population who share a given property or possess certain characteristics (e.g. that of being counterfeit). Suppose now that $n = 10$ units are analysed, so that there are $2^n = 1024$ possible outcomes. The forensic scientist should be able to assign a probability to each of the 1024 possible outcomes. At this point, if it was reasonable to assume that only the observed values $x_1, x_2, \ldots, x_n$ matter and not the order in which they appear, the forensic scientist would have a sensibly simplified task. In fact, the total number of probability assignments would reduce from 1024 to 11, since it is assumed that all sequences are assigned the same probability if they have the same number of 1's, (i.e. successes). This is possible if it is thought that all the items are indistinguishable in the sense that it

does not matter which particular item produced a success (e.g. a positive response) or a failure (e.g. a negative response). Stated otherwise, this means that one's probability assignment is invariant under changes in the order of successes and failures. If the outcomes were permuted in any way, assigned probabilities would be unchanged. For a coin-tossing experiment, Lindley (2014) has expressed this as follows:

> *One way of expressing this is to say that any one toss, with its resulting outcome, may be exchanged for any other with the same outcome, in the sense that the exchange will not alter your belief, expressing the idea that the tosses were done under conditions that you feel were identical. (p. 148)*

The role of exchangeability in the reconciliation of subjective probabilities and frequencies in forensic science is developed in Taroni et al. (2018b). It is possible to give relative frequency an explicit role in probability assignments but this does not mean that probabilities can only be given when relative frequencies are available.

The existence of relative frequencies is not a necessary condition for the assignment of probabilities. Typically, relative frequencies are not available in the case of single (not replicable) events. Other methods of elicitation, such as scoring rules, can be implemented to deal with such situations. An extended discussion on elicitation is given by O'Hagan et al. (2006).

The use of scores for the assessment of forecasts is described in DeGroot and Fienberg (1983). The

association of scores for the assessment of forecasts and the use of scores for the assessment of the performance of methods for evidence evaluation will be made clear later in Section 8.4.3. A score is used to evaluate and compare forecasters who present their predictions of whether or not an event will occur as a subjective probability of the occurrence of that event. A common use for forecasts is that of weather from one day to the next. Let x denote a forecaster's prediction of rain on the following day. Let p be the forecaster's actual subjective probability of rain for that day. Let an arbitrary function $g_1(x)$ be the forecaster's score if rain occurs and let another arbitrary function $g_2(x)$ be their score if rain does not occur. With an assumption that the forecaster wishes to maximise their score, assume that $g_1(x)$ is an increasing function of x and $g_2(x)$ is a decreasing function of x. For a prediction of x and an actual subjective probability of p, the expected score of the forecaster is

$$p\, g_1(x) + (1 - p)g_2(x). \tag{1.3}$$

A *proper scoring rule* is one for which (1.3) is maximised when $x = p$. A *strictly proper scoring rule* is one for which $x = p$ is the *only* value of x that maximises (1.3).

One of the earliest scoring rules, proposed for meteorological forecasts, is the *quadratic scoring rule* (Brier 1950). This score has the property that the forecaster will minimise their subjective expected Brier score on any particular day with

a stated prediction x of their actual subjective probability p of rain on that day. The expected Brier score is then

$$p\,(x-1)^2 + (1-p)\,x^2. \qquad (1.4)$$

This is minimised uniquely when $x = p$. The negative of the Brier score is a strictly proper scoring rule with $g_1(x) = -(x-1)^2$ and $g_2(x) = -x^2$ (minimisation of a function corresponds to maximisation of the negative of the function).

The notion of exchangeability is illustrated with the following example of selection without replacement of items of a particular type, say, Q, from a small population. As an example of what Q might be, consider tablets in a consignment of drugs; the tablets may be either illicit (Q) or licit. The descriptor 'small' for the population size is used to indicate that removal of a member from the population, as in selection without replacement, effects the probability of possession of Q when the next member is selected for removal.

Denote the population size by N. Of the items in the population, R possess Q and $(N-R)$ do not and R is not known. A sample of size $n(< N)$ is taken. The probability the first item selected from the population is of type Q is R/N. If the first member selected from the population possesses Q, the probability the next member selected also possesses Q is $(R-1)/(N-1)$. The population size N is sufficiently small that $(R-1)/(N-1)$ cannot be approximated meaningfully by R/N. Successive draws from the consignment are not independent in that knowledge of the outcome

of one draw affects the probability of a particular outcome at the next draw.

Let X be the number of members of the sample of size n that possess Q. The probability distribution for X is the hypergeometric distribution (Section 4.3.2 and Appendix A.2.5) and

$$\Pr(X = x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}.$$

This distribution does not depend on the order in which the n members are drawn from the population, only on the number x which possess Q and the number $(n - x)$ which do not. The property that the distribution is independent of the order is that of exchangeability.

As R is not known, it is not possible to determine $\Pr(X = x)$. However, it is possible given values for n, N, and x to make inferences about R. A comparison of the frequentist and Bayesian approaches to this small consignment sampling problem is given in Section 4.3.2 and Aitken (1999).

Probabilities based on frequencies may be thought of as objective probabilities. They are considered objective in the sense that there is a well-defined set of circumstances for the long-run repetition of the trials, such that the corresponding probabilities are well-defined and that one's personal or subjective views will not alter the value of the probabilities. Each person considering these circumstances will provide the same values for the probabilities. The frequency model relates to a relative frequency obtained in a

long sequence of trials, assumed to be performed in an identical manner, physically independent of each other. Such a circumstance has certain difficulties. This point of view does not allow a statement of probability for any situation that does not happen to be embedded, at least conceptually, in a long sequence of events giving equally likely outcomes. However, note the following words of Lindley (2004):

> *Objectivity is merely subjectivity when nearly everyone agrees. (p. 87)*

1.7.8 Laws of Probability

There are several laws of probability that describe the values that probability may take and how probabilities may be combined as it has been discussed already in Section 1.7.6. These laws are given here, first for events that are not conditioned on any other information and then for events which are conditioned on other information.

The first law of probability, has already been suggested implicitly.

First Law of Probability

Probability can take any value between 0 and 1, inclusive, and only one of those values. Let R be any event and let $\Pr(R)$ denote the probability that R occurs. Then $0 \leq \Pr(R) \leq 1$. For an event that

is known to be impossible, the probability is zero. Thus if R is impossible, $\Pr(R) = 0$. For an event that is known to be certain, the probability is one. Thus, if R is certain, $\Pr(R) = 1$. This law is sometimes known as the *convexity rule* (Lindley 1991).

Consider the hypothetical example of the balls in the urn of which a proportion b are black and a proportion w white, with no other colours present, such that $b + w = 1$. Proportions lie between 0 and 1; hence $0 \leq b \leq 1$, $0 \leq w \leq 1$. For any event R, $0 \leq \Pr(R) \leq 1$. Consider B, the drawing of a black ball. If there are no black balls in the urn, this event is impossible then $b = 0$. This law is sometimes strengthened to say that a probability can *only* be 0 when the associated event is known to be impossible.

The first law concerns only one event. The next two laws, sometimes known as the second and third laws of probability, are concerned with combinations of events. Events combine in two ways. Let R and S be two events. One form of combination is to consider the event 'R and S', the event that occurs if and only if R and S both occur, sometimes denoted RS. This is known as the *conjunction* of R and S.

Consider the roll of a six-sided fair die. Let R denote the throwing of an odd number. Let S denote the throwing of a number greater than 3 (i.e. a 4, 5, or 6). Then the event 'R and S' denotes the throwing of a 5.

Secondly, consider rolling two six-sided fair die. Let R denote the throwing of a six with the first die. Let S denote the throwing of a six with the second die. Then the event 'R and S' denotes the throwing of a double 6.

The second form of combination is to consider the event 'R or S', the event that occurs if R or S (or both) occurs. This is known as the *disjunction* of R and S.

Consider again the roll of a single six-sided fair die. Let R, the throwing of an odd number (1, 3, or 5), and S, the throwing of a number greater than 3 (4, 5, or 6), be as before. Then 'R or S' denotes the throwing of any number other than a 2 (which is both even and less than 3).

Secondly, consider drawing a card from a well-shuffled pack of 52 playing cards, such that each card is equally likely to be drawn. Let R denote the event that the card drawn is a spade. Let S denote the event that the card drawn is a club. Then the event 'R or S' is the event that the card drawn is from a black suit.

Second Law of Probability

The second law of probability concerns the disjunction 'R or S' of two events. Events are called *mutually exclusive* when the occurrence of one excludes the occurrence of the other. For such events, the conjunction 'R and S' is impossible. Thus $\Pr(R \text{ and } S) = 0$.

If R and S are mutually exclusive events, the probability of their disjunction 'R or S' is equal to the sum of the probabilities of R and S. Thus, for mutually exclusive events,

$$\Pr(R \text{ or } S) = \Pr(R) + \Pr(S). \qquad (1.5)$$

Consider the drawing of a card from a well-shuffled pack of cards with R defined as the drawing of a spade and S the drawing of a club. Then $\Pr(R) = 1/4,\ \Pr(S) = 1/4,\ \Pr(R \text{ and } S) = 0$ (a card may be a spade, a club, neither but not both). Thus, the probability that the card is drawn from a black suit, $\Pr(R \text{ or } S)$ is $1/2$, which equals $\Pr(R) + \Pr(S)$.

Consider the earlier example, the rolling of a single six-sided fair die. Then the events R and S are not mutually exclusive. In the discussion of conjunction it was noted that the event 'R and S' denoted the throwing of a 5, an event with probability $1/6$. The general law, when $\Pr(R \text{ and } S) \neq 0$, is

$$\Pr(R \text{ or } S) = \Pr(R) + \Pr(S) - \Pr(R \text{ and } S).$$

This rule can be easily verified in this case where $\Pr(R) = 1/2,\ \Pr(S) = 1/2,\ \Pr(R \text{ and } S) = 1/6,\ \Pr(R \text{ or } S) = 5/6$.

Before discussing the third law of probability for the conjunction of two events, it is necessary to introduce the ideas of dependence and independence.

1.7.9 Dependent Events and Background Information

Consider, one roll of a fair die with R, the throwing of an odd number as before, and S, the throwing of a number greater than 3, as before. Then, $\Pr(R) = 1/2, \Pr(S) = 1/2, \Pr(R) \times \Pr(S) = 1/4$ but $\Pr(R \text{ and } S) = \Pr(\text{throwing a } 5) = 1/6$. Event R and S are said to be *dependent*.

The third law of probability for dependent events was first presented by Bayes (1763) (see also Barnard 1958; Pearson and Kendall 1970; Poincaré 1912). It is the general law for the conjunction of events. Before the general statement of the third law is made, some discussion of dependence is helpful.

It is useful to consider that a probability assessment depends on two things: the event R whose probability is being considered and the information I available when R is being considered. The probability $\Pr(R \mid I)$ is referred to as a *conditional probability*, acknowledging that R is conditional or dependent on I. Note the use of the vertical bar |. Events listed to the left of it are events whose probability is of interest. Events listed to the right are events whose outcomes are known and which may affect the probability of the events listed to the left of the bar, the vertical bar having the meaning 'given' or 'conditional on'.

Consider a defendant in a trial who may or may not be truly guilty. Denote the event that they are truly guilty by G. The uncertainty associated with their true guilt, the probability that they are truly guilty, may be denoted by $\Pr(G)$. It is a subjective probability. The uncertainty will fluctuate during the course of a trial. It will fluctuate as evidence is presented. It depends on the evidence. Yet neither the notation, $\Pr(G)$, nor the language, the probability of true guilt, makes mention of this dependence. The probability of true guilt at any particular time depends on the knowledge (or information) available at that time. Denote this information by I. It is then possible to speak of the probability of true guilt given, or conditional on, the information I available at that time. This is written as $\Pr(G \mid I)$. If additional evidence E is presented this then becomes, along with I, part of what is known. What is taken as known is then 'E and I', the conjunction of E and I. The revised probability of true guilt is $\Pr(G \mid E \text{ and } I)$. If the information concerns individual S at time t as in Section 1.7.5 the probability can be written as $\Pr(G \mid E, I_{s,t})$.

All probabilities should be thought of as conditional probabilities. Personal experience informs judgements made about events. For example, judgement concerning the probability of rain the following day is conditioned on personal

experiences of rain following days with similar weather patterns to the current one. Similarly, judgement concerning the value of evidence or the guilt of a PoI is conditional on many factors. These include other evidence at the trial but may also include a factor to account for the perceived reliability of the evidence. There may be eyewitness evidence that the PoI was seen at the scene of the crime but this evidence may be felt to be unreliable. Its value will then be lessened.

The value of scientific evidence will be conditioned on the background data relevant to the type of evidence being assessed. Evidence concerning frequencies of different DNA profiles will be conditioned on information regarding ethnicity of the people concerned for the values of these frequencies. Evidence concerning distributions of the refractive indices of glass fragments will be conditioned on information regarding the type of glass from which the fragments have come (e.g. building window, car headlights etc.). The existence of such conditioning events will not always be stated explicitly. However, they should not be forgotten. As stated above, all probabilities may be thought of as conditional probabilities. The first two laws of probability can be stated in the new notation, for events R, S and information I as:

First law of probability for dependent events

$$0 \leq \Pr(R \mid I) \leq 1. \qquad (1.6)$$

If I is known, $\Pr(I \mid I) = 1$ and $\Pr(\text{not } I \mid I) = 0$.

Second law of probability for dependent events

$$\Pr(R \text{ or } S \mid I) = \Pr(R \mid I) + \Pr(S \mid I) - \Pr(R \text{ and } S \mid I). \qquad (1.7)$$

Events R and S are said to be dependent if the knowledge that R has occurred affects the probability that S will occur, and *vice versa*. For example, let R be the outcome of a draw of a card from a well-shuffled pack of 52 playing cards. This card is not replaced in the pack so there are now only 51 cards in the pack. Let S be the draw of a card from this reduced pack of cards. Let R be the event '*an Ace is drawn*'. Thus $\Pr(R) = 4/52 = 1/13$. (Note here the conditioning information I that the pack is well-shuffled, with its implication that each of the 52 cards is equally likely to be drawn has been omitted for simplicity of notation; explicit mention of I will be omitted in many cases but its existence should never be forgotten.) Let S be the event '*an Ace is drawn*' also. Then $\Pr(S \mid R)$ is the probability that an Ace was drawn at the second draw, given that an Ace was drawn at the first draw (and given everything else that is known, in particular that the first card was not replaced). There are 51 cards at the time of the second draw of which 3 are Aces. (Remember that an Ace was drawn the first time which is the information contained in R.) Thus $\Pr(S \mid R) = 3/51$. It is now possible to formulate the third law of probability for dependent events.

Third law of probability for dependent events

$$\Pr(R \text{ and } S \mid I) = \Pr(R \mid I) \times \Pr(S \mid R \text{ and } I). \tag{1.8}$$

Thus in the example of the drawing of the Aces from the pack, the probability of drawing two Aces is

$$\Pr(R \text{ and } S \mid I) = \Pr(R \mid I) \times \Pr(S \mid R \text{ and } I)$$
$$= \frac{4}{52} \times \frac{3}{51}.$$

Example 1.3. A study of the brains of 120 road accident fatalities given in Pittella and Gusmäo (2003, Table 2), reproduced in Lucy (2005) observed the numbers of diffuse vascular injuries (DVI) and diffuse axonal injuries (DAI) with the results presented in Table 1.3.

Denote the presence of DVI by R and the presence of DAI by S. Then various probabilities

Table 1.3 Presence and absence of diffuse vascular injuries (DVI) and diffuse axonal injuries (DAI) in 120 road accident fatalities.

	DAI		
DVI	Present	Absent	Total
Present	14	0	14
Absent	82	24	106
Total	96	24	120

Source: From Pittella and Gusmäo (2003). ©ASTM International. Reprinted with permissions of ASTM International.

for the incidences of the two types of injuries in the population of road accident fatalities can be estimated from this sample of 120 fatalities. Thus $\Pr(R)$ is estimated by 14/120, the total number of DVI divided by the total number of fatalities. Similarly $\Pr(S)$ is estimated by 96/120, the number of DAI divided by the total number of fatalities.

The third law of probability for dependent events (1.8) can be verified using Table 1.3. For example,

$$\frac{14}{120} = \Pr(R \text{ and } S) = \Pr(R) \times \Pr(S \mid R)$$
$$= \frac{14}{120} \times \frac{14}{14}.$$

Alternatively

$$\frac{14}{120} = \Pr(R \text{ and } S) = \Pr(S) \times \Pr(R \mid S)$$
$$= \frac{14}{96} \times \frac{96}{120}.$$

Thus, for dependent events, R and S, the third law of probability, (1.8) may be written as

$$\Pr(R \text{ and } S) = \Pr(S \mid R) \times \Pr(R)$$
$$= \Pr(R \mid S) \times \Pr(S), \qquad (1.9)$$

where the conditioning on I has been omitted.

1.7.9.1 Independence

If two events R and S are such that, given background information I,

$$\Pr(R \mid I) = \Pr(R \mid S, I)$$

they are said to be *independent*. Uncertainty about R is independent of the knowledge of S. From (1.9) it can be seen that

$$\Pr(RS \mid I) = \Pr(R \mid I) \times \Pr(S \mid I).$$

Independent events are exchangeable. It is not necessarily the case that exchangeable events are independent. See Taroni et al. (2018b) for a discussion. Also, two events which are mutually exclusive cannot be independent. As an example of independence, consider the rolling of two six-sided fair dice, A and B say. The outcome of the throw of A does not affect the outcome of the throw of B. If A lands 6 uppermost, this result does not alter the probability that B will land 6 uppermost. The same argument applies if one die is rolled two or more times. Outcomes of earlier throws do not affect the outcomes of later throws. Similarly, with the drawing of two cards from a pack of 52 cards, if the first card drawn is replaced in the pack, and the pack shuffled, before the second draw, the outcomes of the two draws are independent. The probability of drawing two aces is $4/52 \times 4/52$. This can be compared with the probability $4/52 \times 3/51$ if the first card drawn was not replaced.

Third law of probability for independent events

The third law, assuming R and S independent, and conditional on I is

$$\Pr(R \text{ and } S \mid I) = \Pr(R \mid I) \times \Pr(S \mid I). \quad (1.10)$$

Notice that the event I appears as a conditioning event in *all* the probability expressions. The laws are the same as before but with this simple extension.

Consider Table 1.3 again. If DVI and DAI were independent then the probability of both occurring in a road accident fatality would be the product of the probability of each happening separately. Thus

$$\begin{aligned}\Pr(R \text{ and } S) &= \Pr(R) \times \Pr(S) = \frac{14}{120} \times \frac{96}{120} \\ &= 0.094,\end{aligned}$$

However, it is not the case that 9.4% of road accident fatalities have both injuries. An examination of Table 1.3 illustrates that this is not so. From Table 1.3 it can be seen that $14/120 = 0.12$ or 12% of fatalities have both injuries. In such a situation where $\Pr(R \text{ and } S) \neq \Pr(R) \times \Pr(S)$ it can be said that DVI and DAI are not independent.

As another example of the use of the ideas of independence, consider a diallelic system in genetics in which the alleles are denoted A and a, with $\Pr(A) = p, \Pr(a) = q; \Pr(A) + \Pr(a) = p + q = 1$. This gives rise to three genotypes that, assuming Hardy–Weinberg equilibrium to hold, are expected to have the following probabilities

- p^2 (homozygotes for allele A),
- $2pq$ (heterozygotes),
- q^2 (homozygotes for allele a).

The genotype probabilities are calculated by simply multiplying the two allele probabilities together on the assumption that the allele inherited from one's father is independent of the allele inherited from one's mother. The factor 2 arises in the heterozygous case because two cases must be considered, that in which allele A was contributed by the mother and allele a by the father, and *vice versa*. Both of these cases have probability pq because of the assumption of independence (see Table 1.4). Note that $p^2 + 2pq + q^2 = (p + q)^2 = 1$. The particular locus under consideration is said to be in Hardy–Weinberg equilibrium when the two parental alleles are considered as independent.

This law may be generalised to more than two events. Consider n events $S_1, S_2, \ldots, S_n$. If they are mutually independent then

$$\Pr(S_1 \text{ and } S_2 \text{ and } \ldots \text{ and } S_n) =$$

$$\Pr(S_1) \times \Pr(S_2) \times \cdots \times \Pr(S_n) = \prod_{i=1}^{n} \Pr(S_i).$$

Table 1.4 Genotype probabilities, assuming Hardy–Weinberg equilibrium, for a diallelic system with allele probabilities p and q.

Allele from mother	Allele from father	
	A (p)	a (q)
A (p)	p^2	pq
a (q)	pq	q^2

1.7.10 Law of Total Probability

Events $S_1, S_2, \ldots, S_n$ are said to be *mutually exclusive and exhaustive* if one of them has to be true and only one of them can be true; they exhaust the possibilities and the occurrence of one excludes the possibility of any other. Alternatively, they are called a *partition*. The event (S_1 or ... or S_n) formed from the conjunction of the individual events $S_1, \ldots, S_n$ is certain to happen since the events are exhaustive and exclusive. Thus, it has probability 1 and

$$\Pr(S_1 \text{ or } \ldots \text{ or } S_n) = \Pr(S_1) + \cdots + \Pr(S_n) = 1, \tag{1.11}$$

a generalisation of the second law of probability, (1.7), for exclusive events. Consider as an example allelic distributions at a locus, e.g. locus *TPOX*. There are five alleles, 8, 9, 10, 11, and 12, and these are mutually exclusive and exhaustive.

Consider $n = 2$ for events S_1 and S_2. Let R be any other event. The events 'R and S_1' and 'R and S_2' are exclusive. They cannot both occur. The event "'R and S_1' *or* 'R and S_2' " is simply R. For example, let S_1 be male, S_2 be female, R be left-handed. Then

- 'R and S_1' denotes a left-handed male,
- 'R and S_2' denotes a left-handed female.

The event ' "R and S_1" or "R and S_2" ' is the event that a person is a left-handed male or a left-handed female, which implies the person is left-handed (R).

Thus,

$$\begin{aligned}\Pr(R) &= \Pr(R \text{ and } S_1) + \Pr(R \text{ and } S_2) \\ &= \Pr(R \mid S_1)\Pr(S_1) + \Pr(R \mid S_2)\Pr(S_2).\end{aligned}$$

The argument extends to any number of mutually exclusive and exhaustive events to give the law of total probability.

Law of Total Probability

If $S_1, S_2, \ldots, S_n$ are n mutually exclusive and exhaustive events,

$$\Pr(R) = \Pr(R \mid S_1)\Pr(S_1) + \cdots + \Pr(R \mid S_n)\Pr(S_n). \tag{1.12}$$

This is sometimes known as the *extension of the conversation* (Lindley 1991)

An example for blood types and paternity cases is given by Lindley (1991). Consider two possible groups, S_1 (Rh−) and S_2 (Rh+) for the father, so here $n = 2$. Assume the relative frequencies of the two groups are p and $(1 - p)$, respectively. The child is Rh− (event R) and the mother is also Rh− (event M). The probability of interest is the probability a Rh− mother will have a Rh− child, in symbols $\Pr(R \mid M)$. This probability is not easily derived directly but the derivation is fairly straightforward if the law of total probability is invoked to include the father.

$$\begin{aligned}\Pr(R \mid M) = {} & \Pr(R \mid M \text{ and } S_1)\Pr(S_1 \mid M) \\ & + \Pr(R \mid M \text{ and } S_2)\Pr(S_2 \mid M).\end{aligned} \tag{1.13}$$

This is a generalisation of the law to include information M. If both parents are Rh−, event (M and S_1), then the child is Rh− with probability 1, so $\Pr(R \mid M \text{ and } S_1) = 1$. If the father is Rh+ (the mother is still Rh−), event S_2, then $\Pr(R \mid M \text{ and } S_2) = 1/2$. Assume that parents mate at random with respect to the Rhesus quality. Then $\Pr(S_1 \mid M) = p$, the relative frequency of Rh− in the population, independent of M. Similarly, $\Pr(S_2 \mid M) = 1 - p$, the relative frequency of Rh+ in the population. These probabilities can now be inserted in (1.13) to obtain

$$\Pr(R \mid M) = 1(p) + \frac{1}{2}(1 - p) = (1 + p)/2,$$

for the probability that a Rh− mother will have a Rh− child. This result is not intuitively obvious, unless one considers the approach based on the law of total probability.

An example using DNA profiles is given in Evett and Weir (1998). According to the 1991 census, the New Zealand (NZ) population consists of 83.47% Caucasians, 12.19% Maoris, and 4.34% Pacific Islanders; denote the event that a person chosen at random from the 1991 NZ population is Caucasian, Maori, or Pacific Islander as *Ca*, *Ma*, and *Pa*, respectively. The probabilities of finding the same YNH24 genotype g (event G) in a crime sample for a Caucasian, Maori, or Pacific Islander are 0.012, 0.045, and 0.039, respectively. These values are the assessments for the following three conditional probabilities: $\Pr(G \mid \text{Ca})$, $\Pr(G \mid \text{Ma})$,

$\Pr(G \mid \text{Pa})$. Then the probability of finding the YNH24 genotype, G, in a person taken at random from the whole population of New Zealand is

$$\begin{aligned}\Pr(G) &= \Pr(G \mid \text{Ca})\Pr(\text{Ca}) + \Pr(G \mid \text{Ma})\Pr(\text{Ma}) \\ &\quad + \Pr(G \mid \text{Pa})\Pr(\text{Pa}) \\ &= 0.012 \times 0.8347 + 0.045 \times 0.1219 \\ &\quad + 0.039 \times 0.0434 \\ &= 0.017.\end{aligned}$$

A further extension of this law to consider probabilities for combinations of genetic marker systems in a racially heterogeneous population has been given by Walsh and Buckleton (1988). Let C and D be two genetic marker systems with realisations C and $\bar{C}$, D, and $\bar{D}$. Let S_1 and S_2 be two mutually exclusive and exhaustive subpopulations such that a person from the population belongs to one and only one of S_1 and S_2. Let $\Pr(S_1)$ and $\Pr(S_2)$ be the probabilities that a person chosen at random from the population belongs to S_1 and to S_2, respectively. Then $\Pr(S_1) + \Pr(S_2) = 1$. Within each subpopulation C and D are independent so that the probability an individual chosen at random from one of these subpopulations is of type CD is simply the product of the individual probabilities. Thus

$$\begin{aligned}\Pr(CD \mid S_1) &= \Pr(C \mid S_1) \times \Pr(D \mid S_1), \\ \Pr(CD \mid S_2) &= \Pr(C \mid S_2) \times \Pr(D \mid S_2).\end{aligned}$$

However, such a so-called *conditional independence* result does not imply unconditional independence (i.e. that $\Pr(CD) = \Pr(C) \times \Pr(D)$). The probability that an individual chosen at random from the population is *CD*, without regard to his subpopulation membership, may be written as follows

$$\begin{aligned}\Pr(CD) &= \Pr(CDS_1) + \Pr(CDS_2)\\ &= \Pr(CD \mid S_1) \times \Pr(S_1)\\ &\quad + \Pr(CD \mid S_2) \times \Pr(S_2)\\ &= \Pr(C \mid S_1) \times \Pr(D \mid S_1) \times \Pr(S_1)\\ &\quad + \Pr(C \mid S_2) \times \Pr(D \mid S_2) \times \Pr(S_2).\end{aligned}$$

This is not necessarily equal to $\Pr(C) \times \Pr(D)$ as is illustrated in the following example. Let $\Pr(C \mid S_1) = \gamma_1$, $\Pr(C \mid S_2) = \gamma_2$, $\Pr(D \mid S_1) = \delta_1$, $\Pr(D \mid S_2) = \delta_2$, $\Pr(S_1) = \theta$, and $\Pr(S_2) = 1 - \theta$. Then

$$\begin{aligned}\Pr(CD) &= \gamma_1\delta_1\theta + \gamma_2\delta_2(1-\theta),\\ \Pr(C) &= \gamma_1\theta + \gamma_2(1-\theta),\\ \Pr(D) &= \delta_1\theta + \delta_2(1-\theta).\end{aligned}$$

The product of $\Pr(C)$ and $\Pr(D)$ is not necessarily equal to $\Pr(CD)$. Suppose, for example that $\theta = 0.40$, $\gamma_1 = 0.10$, $\gamma_2 = 0.20$, $\delta_1 = 0.15$, and $\delta_2 = 0.05$. Then

$$\begin{aligned}\Pr(CD) &= \gamma_1\delta_1\theta + \gamma_2\delta_2(1-\theta)\\ &= 0.10 \times 0.15 \times 0.4 + 0.20 \times 0.05 \times 0.6\\ &= 0.0120.\end{aligned}$$

$$\Pr(C) = \gamma_1\theta + \gamma_2(1-\theta) = 0.04 + 0.12 = 0.16.$$

$$\Pr(D) = \delta_1\theta + \delta_2(1-\theta) = 0.06 + 0.03 = 0.09.$$

$$\Pr(C) \times \Pr(D) = 0.0144 \neq 0.0120 = \Pr(CD).$$

1.7.11 Updating of Probabilities

Notice that the probability of true guilt is a subjective probability, as mentioned before (Section 1.7.4). Its value will change as evidence accumulates. Also, different people will have different values for it. The following examples, adapted from similar ones in DeGroot (1970), illustrate how probabilities may change with increasing information. The examples have several parts and each part has to be considered in turn without information from a later part.

Example 1.4.

(a) Consider four events S_1, S_2, S_3, and S_4. Event S_1 is that the area of Lithuania is no more than 50 000 km^2, S_2 is the event that the area of Lithuania is greater than 50 000 km^2 but no more than 75 000 km^2, S_3 is the event that the area of Lithuania is greater than 75 000 km^2 but no more than 100 000 km^2, and S_4 is the event that the area of Lithuania is greater than 100 000 km^2. Assign probabilities to each of these four events. Remember that these are four mutually exclusive events and

that the four probabilities should add up to 1. Which do you consider the most probable and what probability do you assign to it? Which do you consider the least probable and what probability do you assign to it?

(b) Now, consider the information that Lithuania is the 25th largest country in Europe (excluding Russia). Use this information to reconsider your probabilities in part (a).

(c) Consider the information that Estonia, which is the 30th largest country in Europe, has an area of 45 000 km^2, and use it to reconsider your probabilities from the previous part.

(d) Consider the information that Austria, which is the 21st largest country in Europe has an area of 84 000 km^2, and use it to reconsider your probabilities from the previous part.

The area of Lithuania is given at the end of the chapter.

Example 1.5.

(a) Imagine you are on a jury. The trial is about to begin but no evidence has been led. Consider the two events: H_p the defendant is truly guilty; and H_d, the defendant is innocent. What are your probabilities for these two events?

(b) The defendant is a tall Caucasian male. An eyewitness says he saw a tall Caucasian male running from the scene of the crime. What are your probabilities now for H_p and H_d?

(c) A bloodstain at the scene of the crime was identified as coming from the criminal. A partial DNA profile has been obtained, with proportion 2% in the local Caucasian population. What are your probabilities now for H_p and H_d?

(d) A window was broken during the commission of the crime. Fragments of glass were found on the defendant's clothing of a similar refractive index to that of the crime window. What are your probabilities now for H_p and H_d?

(e) The defendant works as a demolition worker near to the crime scene. Windows on the demolition site have refractive indices similar to the crime window. What are your probabilities now for H_p and H_d?

This example is designed to mimic the presentation of evidence in a court case. Part (a) asks for a prior probability of guilt before the presentation of any evidence. It may be considered as a question concerning the understanding of the dictum 'innocent until proven guilty'. See Section 2.7 for further discussion of this with particular reference to the logical problem created if a prior probability of zero is assigned to the event that the suspect is guilty.

Part (b) involves two parts. First, the value of the similarity in physical characteristics between the defendant and the person running from the scene of the crime, assuming the eyewitness is reliable, has to be assessed. Secondly, the

assumption that the eyewitness is reliable has to be assessed.

In part (c) it is necessary to check that the defendant has the same profile. It is not stated that he has but if he has not he should never have been a defendant. Secondly, is the local Caucasian population the relevant population? The evaluation of evidence of the form in (c) is discussed in Chapter 5.

The evaluation of refractive index measurements mentioned in (d) is discussed in Chapter 7. Variation both within and between windows has to be considered. Finally, how information about the defendant's lifestyle may be considered is discussed in Chapter 6.

It should be noted that the questions asked initially in Example 1.5 are questions that should be addressed by the judge and/or jury. The forensic scientist is concerned with the evaluation of their evidence, not with probabilities of guilt or innocence. These probabilities are the concern of the jury. The jury combines the evidence of the scientist with all other evidence and uses its judgement to reach a verdict. The theme of this book is the evaluation of evidence. Discussion of issues relating to guilt or otherwise of PoIs will not be very detailed.

As a tail piece to this chapter, the area of Lithuania is $65\ 301\ \text{km}^2$.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

2

The Evaluation of Evidence

The evaluation of evidence is not to be confused with the interpretation of evidence. The two terms *evaluation* and *interpretation* are sometimes considered as synonyms but it is helpful to think of them as different from each other. 'Evaluation' concentrates on the derivation of a value for the evidence, in a way to be described in this chapter. 'Interpretation' refers to the meaning attached to such a value in the case as a whole.

2.1 ODDS

2.1.1 Complementary Events

There is a measure of uncertainty, known as *odds* (also known as *betting quotients*), which will be familiar to people who know about gambling. Bookmakers quote odds in sporting events such

as horse races or football matches. It has been mentioned briefly in Section 1.7.6. For example, a particular horse may be given odds of '6 to 1 against' it winning a race or a particular football team may be quoted at odds of '3 to 2 on' to win a match or, equivalently, '3 to 2 in favour' of winning the match. Odds are equivalent to probability. The aforementioned phrases can be related directly to probability statements about the probability of the horse winning its race or the football team winning its match.

First, an event, known as the negation, or *complement*, of another event has to be introduced and given some notation. Let R be an event. Then the negation or complement of R is the event that is true when R is false and false when R is true. It is denoted $\bar{R}$ and read as '*R-bar*'. The events R and $\bar{R}$ are known as *complementary events*.

The union of two or more events is known as a *disjunction*. Let $E_i, i = 1, \ldots, 6$ be the event that denotes the throw of an i in a six-sided die. Then the disjunction of E_1, E_3, and E_5 is the throw of an odd number. The occurrence of two or more events simultaneously is known as a *conjunction*. Thus the conjunction of the event 'throw an odd number in a six-sided die' and 'throw a number less than 4' is the event that a one or three is thrown. For human characteristics, consider the events a person is blue-eyed and a person has red hair. The conjunction of these two events are people with blue eyes and red hair. The disjunction

of these two events are people with blue eyes or with red hair or with both. Note the use of the word 'and' for conjunction and 'or' for disjunction.

Often in this book comparison will be made of the probability of the evidence under two competing propositions, that put forward by the prosecutor and that put forward by the defence. The proposition put forward by the prosecution will be denoted H_p. The proposition put forward by the defence will be denoted by H_d. The subscripts p and d denote *prosecution* and *defence*, respectively. The letter H denotes hypothesis and the letter has stuck despite the more common usage now of the term *proposition*. Note that, from now on, the term *proposition* is used in preference to the word *hypothesis* to designate the form of words deemed to be relevant for the scenario under study. Also, and for the sake of simplicity, apostrophes (' ') will be omitted when describing verbally the content of a proposition, except when a proposition is part of a sentence.

Propositions may be complementary in the same way as events are said to be complementary. One and only one can be true. They are mutually exclusive. The propositions need not be exhaustive; they need not be chosen to cover all possible explanations for the evidence, no matter how outlandish. The two propositions may denote complementary events, such as (truly) *Guilty* and (truly) *Not guilty*. However, there will be occasions on which the events denoted are not complementary, such as

'*Person of interest*[1] *A and one unknown person were present at the crime scene*' and '*Two unknown people were present at the crime scene*'. There are many other events not covered by these two propositions, such as the naming of the two individuals at the crime scene or the consideration of less than, or more than, two people at the crime scene.

2.1.2 Examples

(1) A coin is tossed. Let R be the event it lands heads. Then $\bar{R}$ is the event that it lands tails. If the coin is fair, $\Pr(R) = 1/2, \Pr(\bar{R}) = 1/2$.

(2) A six-sided die is rolled. Let R be the event that a six is face uppermost. Then $\bar{R}$ is the event that a 1, 2, 3, 4, or 5 is rolled. If the die is fair, $\Pr(R) = 1/6, \Pr(\bar{R}) = 5/6$.

(3) A pill is checked to see if it is licit or illicit. Let R be the event that it is illicit. Then $\bar{R}$ is the event that it is licit.

(4) A person is charged with a crime. Let G be the event that he is truly guilty, not just found guilty by a jury. Then $\bar{G}$ is the event that he is truly not guilty.

Notice that the event 'R or $\bar{R}$', formed from the disjunction of R and its complement $\bar{R}$, is certain. Thus it has probability 1. Also, since R and $\bar{R}$ are

[1]Note the use of the phrase 'person of interest' (PoI); recall again (see Section 1.1) that this phrase is used instead of the word 'suspect' as it is felt to represent more accurately the status of the person concerned in an investigation.

mutually exclusive,

$$1 = \Pr(R \text{ or } \bar{R}) = \Pr(R) + \Pr(\bar{R})$$

and hence

$$\Pr(\bar{R}) = 1 - \Pr(R). \tag{2.1}$$

In general, for complementary events R and $\bar{R}$,

$$\Pr(R) + \Pr(\bar{R}) = 1. \tag{2.2}$$

It is now possible to define odds.

2.1.3 Definition of Odds

If an event R has probability $\Pr(R)$ of occurring, the *odds* against R are

$$\Pr(\bar{R}) / \Pr(R).$$

From (2.1), the odds against R are

$$\frac{1 - \Pr(R)}{\Pr(R)}. \tag{2.3}$$

The odds in favour of R are

$$\frac{\Pr(R)}{1 - \Pr(R)}.$$

Given a probability for an event, it is possible to derive the odds against the event.

Given a value for the odds against an event, it is possible to determine the probability of that event occurring. Thus, if the horse has odds of 6 to 1

against it winning the race and R is the event that it wins the race, then

$$\frac{1 - \Pr(R)}{\Pr(R)} = 6,$$

where '6 to 1' is taken as the ratio 6/1 and written as 6. Then

$$\begin{aligned} 1 - \Pr(R) &= 6 \times \Pr(R) \\ 1 &= \{6 \times \Pr(R)\} + \Pr(R) \\ &= 7 \times \Pr(R). \end{aligned}$$

Thus $\Pr(R) = 1/7$.

The phrases 'odds on' and 'odds in favour of' are equivalent and are used as the reciprocal of 'odds against'. Consider the football team that is '3 to 2 on' to win its match. The phrase '3 to 2' is taken as the ratio 3/2 as this is odds *on*. The relationship between odds and probability is written as

$$\frac{\Pr(R)}{1 - \Pr(R)} = \frac{3}{2}.$$

Thus

$$\begin{aligned} 2 \times \Pr(R) &= 3\{1 - \Pr(R)\} \\ 5 \times \Pr(R) &= 3 \\ \Pr(R) &= 3/5. \end{aligned}$$

The general result may be derived as follows. Let O denote the odds against the occurrence of an event R. Then

$$O = \frac{1 - \Pr(R)}{\Pr(R)},$$

$$O \times \Pr(R) = 1 - \Pr(R),$$
$$(O+1) \times \Pr(R) = 1,$$
$$\Pr(R) = \frac{1}{O+1}.$$

This can be verified directly for the horse whose odds were 6 to 1 against it winning (with $O = 6$). For the football team with odds of 3 to 2 on this can be taken as 2 to 3 against ($O = 2/3$) and the result follows. Odds equal to 1 are known as *evens*.

The concept of odds is an important one in the evaluation of evidence. Evidence is evaluated for its effect on the probability of a certain supposition about a PoI (before they come to trial) or defendant (whilst a trial is in progress). This supposition may be that the PoI was present at the crime scene, a *source proposition* (see Chapter 5). It is this supposition that will be most discussed in this book. Initially, however, the discussion will be in terms of the effect of evidence on the probabilities of the guilt (H_p) and the innocence (H_d) of a suspect. These are two complementary events. The ratio of the probabilities of these two events, $\Pr(H_p)/\Pr(H_d)$, is the odds against innocence or the odds in favour of guilt. Notice, also, that the events are that the suspect is truly guilty or truly innocent, not that he is judged to be guilty or innocent. The same principles concerning odds also apply for conditional probabilities. Given background information I, the ratio $\Pr(H_p \mid I)/\Pr(H_d \mid I)$ is the odds in

favour of guilt, given I. Much of this book will be concerned with the effect on the odds in favour of a supposition about the PoI of the evidence E under consideration.

There are occasions when the prosecution and defence proposition are not complementary. In such instances it is not possible to determine $\Pr(H_p)$ or $\Pr(H_d)$ from the odds, only the effect the statistic known as the likelihood ratio (LR, Section 2.3) has on the odds. Note that in this context the term 'odds' is a misnomer as the term strictly applies to the ratio of the probability of complementary events.

2.2 BAYES' THEOREM

Bayes' theorem is an important part of the process of the consideration of the odds. In fact, the theorem permits the revision based on new information of a measure of uncertainty about the truth or otherwise of an outcome or issue (such as a hypothesis or proposition). This perspective is common to numerate scientific fields where data are combined with prior or background information to give posterior probabilities for a particular outcome or issues. An essential feature of Bayesian inference is that it permits the move from prior (initial or pre-test) to posterior (final or post-test) probabilities on the basis of data.

2.2.1 Statement of the Theorem

Consider the last two parts of the third law of probability as given in (1.9), namely, that for events R and S,

$$\Pr(S \mid R) \times \Pr(R) = \Pr(R \mid S) \times \Pr(S).$$

If $\Pr(R) \neq 0$, it is possible to divide by $\Pr(R)$ and obtain the *Bayes' theorem* for two events, R and S,

$$\Pr(S \mid R) = \frac{\Pr(R \mid S) \times \Pr(S)}{\Pr(R)} \qquad (2.4)$$

assuming $\Pr(R) \neq 0$.

2.2.2 Examples

An important example of such reasoning is found in medical diagnosis. Consider the following example where a doctor in a clinic is interested in the proposition 'This patient has disease S'. By regarding the patient as a random member of a large collection (population) of patients presenting themselves in the clinic (for a discussion on the population characteristics relevant to a case, see Section 1.7.11), the doctor associates a probability with the proposition of interest: this probability is the prior (or pre-test) probability the patient has S. Note that such a probability is always conditioned on background information

about the patient (i.e. age, gender, medical history...).

It is the probability $\Pr(S)$ that a person has the disease S, before any test results or new observations are taken. Suppose the doctor then carries out a test (e.g. a blood test) that gives a positive result; call this event R. After that, the doctor is now interested in assessing the new probability that the patient has disease S. This new value is the posterior or post-test probability $\Pr(S \mid R)$ because it refers to a new situation, as expressed by (2.4).

The probability of a positive blood test can be expanded using the *extension of the conversation* (Section 1.7.14). A positive blood test result could be considered under two competing situations: first, the patient has disease S (event S), and second, the patient does not have disease S (event $\bar{S}$). So $\Pr(R)$ becomes

$$\Pr(R \mid S) \times \Pr(S) + \Pr(R \mid \bar{S}) \times \Pr(\bar{S}),$$

and the posterior probability

$$\Pr(S \mid R) = \frac{\Pr(R \mid S) \times \Pr(S)}{\Pr(R \mid S) \times \Pr(S) + \Pr(R \mid \bar{S}) \times \Pr(\bar{S})}. \tag{2.5}$$

A numerical verification of this result, though not in the context of a disease, is available from Table 2.5. Let R denote Female (F), S denote a plain arch in a fingerprint (PA). Then $\Pr(R \mid S) = 57/185, \Pr(S) = 185/207, \Pr(R) = 62/207$, and

$$\frac{\Pr(R \mid S) \times \Pr(S)}{\Pr(R)} = \frac{\frac{57}{185} \times \frac{185}{207}}{\frac{62}{207}} = \frac{57}{62} = \Pr(S \mid R).$$

The importance of Bayes' theorem is that it links $\Pr(S)$ with $\Pr(S \mid R)$. The uncertainty about S as given by $\Pr(S)$ on the right-hand side of (2.4) is updated by the knowledge about R to give the uncertainty about S as given by $\Pr(S \mid R)$ on the left-hand side of (2.4). Note that the connection between $\Pr(S)$ and $\Pr(S \mid R)$ involves both $\Pr(R \mid S)$ and $\Pr(R)$.

Reconsider the previous simple example where a doctor is interested in the probability the patient has disease S given the positive blood result R. For the quantitative assessment of conditional probabilities involving test results, it is important that in an earlier stage (that is before consideration of a particular patient), the blood test used by the doctor is evaluated using two groups of patients with and without the disease. The groups are classified using a reference test (a so-called *gold standard*) to obtain a two-by-two table, Table 2.1, known as a *contingency table*. There are n patients in total. The number of patients in each category is identified by the subscripts. The sums of pairs of numbers in rows or columns in the body of the table are the values in the margins (bottom row and right-hand column). Thus, for example, $n_{RS} + n_{R\bar{S}} = n_R$, the number of patients with a positive blood test. In medical terminology it is common to refer to the *sensitivity* and *specificity* of a test. Sensitivity is the probability of a positive result in the blood test given that the patient has disease S. It is estimated by the ratio of n_{RS} to n_S, the proportion of positive patients in the diseased

Table 2.1 Two-by-two contingency table for frequencies for the tabulation of patients with or without a disease (S or $\bar{S}$) and a blood test positive or negative (R or $\bar{R}$).

	S	$\bar{S}$	Total
R	n_{RS}	$n_{R\bar{S}}$	n_R
$\bar{R}$	$n_{\bar{R}S}$	$n_{\bar{R}\bar{S}}$	$n_{\bar{R}}$
Total	n_S	$n_{\bar{S}}$	n

group. Specificity is the probability of a negative result in the blood test given that the patient does not have the disease. It is estimated by the ratio of $n_{\bar{R}\bar{S}}$ to $n_{\bar{S}}$, the proportion of negative patients in the non-diseased group. Sensitivity and specificity provide a measure of the quality of a test, with high values implying high quality. Note that values in Table 2.1 refer to a sample of patients and implicitly assume that the parameters of interest (the proportions called *sensitivity* and *specificity*) are directly observable from the sample. However, the proportions in a relevant population are unknown. The connection between the sample proportion and the population proportion is presented in Section 4.2.

Thus, in (2.5) $\Pr(S)$ represents the prior probability that the patient has disease S (in medical terms, this probability is also called *prevalence*), $\Pr(\bar{S})$ equals $1 - \Pr(S)$. $\Pr(R \mid S)$ is the sensitivity of the test. $\Pr(\bar{R} \mid \bar{S})$ is the specificity of the test. $\Pr(R \mid \bar{S})$, which is $1 - \Pr(\bar{R} \mid \bar{S})$, is known as the

false positive rate and is estimated by the ratio of $n_{R\bar{S}}$ to $n_{\bar{S}}$. The *false negative* rate is $\Pr(\bar{R} \mid S)$ and is estimated by the ratio of $n_{\bar{R}S}$ to n_S.

Table 2.1 can also be presented using probabilities instead of frequencies (see, for example, Leonard (2000)).

In Table 2.2, $\Pr(S)$ is the prior probability or prevalence of the disease in the relevant population and $\Pr(S, R) = \Pr(R \mid S) \times \Pr(S)$ is assessed using the sensitivity of the test. $\Pr(S, \bar{R})$ may then be calculated by subtraction, $\Pr(S) - \Pr(S, R)$. An analogous procedure is adopted for the column $\bar{S}$.

The distinction between $\Pr(S \mid R)$ and $\Pr(R \mid S)$ is very important and needs to be recognised. In $\Pr(S \mid R)$, R is known or given, S is uncertain. In the medical example, the result of the blood test is known, the disease status is unknown. In $\Pr(R \mid S)$, S is known or given, R is uncertain. In the medical example, the disease status is known, the result of the blood test is uncertain.

Table 2.2 Two-by-two contingency table for probabilities for the tabulation of patients with or without a disease (S or $\bar{S}$) and a blood test positive or negative (R or $\bar{R}$).

	S	$\bar{S}$	Total
R	$\Pr(S, R)$	$\Pr(\bar{S}, R)$	$\Pr(R)$
$\bar{R}$	$\Pr(S, \bar{R})$	$\Pr(\bar{S}, \bar{R})$	$\Pr(\bar{R})$
Total	$\Pr(S)$	$\Pr(\bar{S})$	1

Further examples will emphasise the difference between these two conditional probabilities.

Example 2.1. Consider the previous medical diagnosis description illustrated through artificial data presented in Table 2.3, where 100 patients with S and 100 patients without S are chosen. The prior probability $\Pr(S)$ or prevalence of the disease in the relevant population is known to be 0.1.

A medical blood test detects certain symptoms of a disease or condition. Unfortunately, the test may not always register the symptoms when they are present, or it may register them when they are absent. Therefore there is the need of numbers to describe the accuracy of the test; these are the sensitivity and the specificity of the performed test. The sensitivity is the probability that a person with the disease is correctly diagnosed. The specificity is the probability that a disease-free individual is correctly diagnosed. Letting S be the event that a person has the disease and R stand for the event

Table 2.3 Two-by-two contingency table for artificial frequencies for the tabulation of patients with or without a disease (S or $\bar{S}$) given a blood test positive or negative (R or $\bar{R}$).

	S	$\bar{S}$	Total
R	95	1	96
$\bar{R}$	5	99	104
Total	100	100	200

that the test indicates a positive result. $\Pr(R \mid S)$ and $\Pr(\bar{R} \mid \bar{S})$ stand for the sensitivity and the specificity of the test, respectively.

The sensitivity of the test is estimated by the proportion $95/100$, the specificity of the test is estimated by the proportion $99/100$; $\Pr(R \mid S) = 0.95$ and $\Pr(\bar{R} \mid \bar{S}) = 0.99$.

The posterior probability, $\Pr(S \mid R)$, a given member of the relevant population has the disease given the observation of the symptoms becomes

$$\begin{aligned}\Pr(S \mid R) &= \frac{\Pr(R \mid S) \times \Pr(S)}{\Pr(R \mid S) \times \Pr(S) + \Pr(R \mid \bar{S}) \times \Pr(\bar{S})} \\ &= \frac{0.95 \times 0.1}{0.95 \times 0.1 + (1 - 0.99) \times 0.9} \\ &= 0.913.\end{aligned}$$

If the blood test is positive the probability of having the disease increases from a prior probability of 0.1 to a posterior probability greater than 0.91. Section 4.2.3 develops the same example by considering the connection between the sample proportion and the population proportion.

Example 2.2. First, let S be the event 'I have two arms and two legs' and let R be the event 'I am a monkey'. Then $\Pr(S \mid R) = 1$,[2] whereas $\Pr(R \mid S) \neq 1$. The first probability is equivalent to saying that 'If I am a monkey then I have two arms

[2]It is sensible to set this probability to 1, omitting cases in which malformations are considered.

and two legs'. The second probability is equivalent to saying that 'If I have two arms and two legs, I need not be a monkey'. Similarly, in the previous medical example, a patient is more interested in the probability of not having the disease, given that the test has a positive result, than in the probability of a positive test given that they do not have the disease. The latter probability is the false positive rate, $\Pr(R \mid \bar{S})$, the former is a posterior probability, $\Pr(\bar{S} \mid R)$. For a discussion on this very important point, see Thompson and Schumann (1987) and Saks and Koehler (1991).

Example 2.3. This example is from Lindley (1991). Consider the following two statements.

(1) The death rate last month amongst men is twice that amongst women.

(2) In the deaths registered last month, there were twice as many men as women.

Let M denote male, F denote female, so that M and F are complementary events and the relationship can be written as $(M \equiv \bar{F}, F \equiv \bar{M})$. Let D denote the event of death. Then statements (1) and (2) may be written as

(1) $\Pr(D \mid M) = 2 \Pr(D \mid F)$,

(2) $\Pr(M \mid D) = 2 \Pr(F \mid D)$.

Notice also that $\Pr(M \mid D) + \Pr(F \mid D) = 1$ since M and F are complementary events. Equation (2.2) generalises to include a conditioning event (D in

Table 2.4 Hypothetical results for deaths amongst a population.

	Male	Female	Total
Dead	2	1	3
Alive	98	99	197
Total	100	100	200

this case). Thus from statement (2),

$$1 - \Pr(F \mid D) = 2 \Pr(F \mid D)$$

and

$$\Pr(F \mid D) = 1/3,\ \Pr(M \mid D) = 2/3.$$

It is not possible to make any similar inferences from statement (1) since in that statement it is the conditioning event that alters, not the uncertain event. Table 2.4 illustrates the point numerically.

There are 100 males of whom 2 died, and 100 females of whom 1 died. Thus $\Pr(D \mid M) = 0.02, \Pr(D \mid F) = 0.01$, satisfying (1). There were 3 deaths in total, of whom 2 were male and 1 female, satisfying the previous statement (2).

Example 2.4. Consider the problem of determining which of three sub-populations (Ψ_1, Ψ_2, Ψ_3) an individual belongs to, based on observations of genotypes at several loci and knowledge of genotype relative frequencies in each of the sub-populations (Shoemaker et al., 1999).

The context may be that of a bloodstain found at a crime scene and the question is to determine which of three populations, e.g. Caucasian, Maori, or Western Polynesian, the contributor of the stain belongs (assuming that attention can be restricted to these three sub-populations).

The relevant New Zealand census reported that the population in the country had the following composition: 81.9% Caucasian, 13.7% Maori, and 4.4% Western Polynesian. The probability of the observed genotypes (a DNA forensic profile) X of the individual can be calculated. For this example, suppose the three probabilities $\Pr(X \mid \Psi_1)$, $\Pr(X \mid \Psi_2)$, $\Pr(X \mid \Psi_3)$ have been assigned as $3.96 \times 10^{-9}, 1.18 \times 10^{-8}, 1.91 \times 10^{-7}$, respectively. The prior probabilities $\Pr(\Psi_i)$ for the three sub-populations are 0.819, 0.137, and 0.044. Then

$$\begin{aligned}\Pr(\Psi_1 \mid X) &= \frac{\Pr(X \mid \Psi_1) \cdot \Pr(\Psi_1)}{\Pr(X \mid \Psi_1) \cdot \Pr(\Psi_1) + \Pr(X \mid \Psi_2) \cdot \Pr(\Psi_2) + \Pr(X \mid \Psi_3) \cdot \Pr(\Psi_3)} \\ &= \frac{3.96 \times 10^{-9} \times 0.819}{3.96 \times 10^{-9} \times 0.819 + 1.18 \times 10^{-8} \times 0.137 + 1.91 \times 10^{-7} \times 0.044} \\ &= 0.245,\end{aligned}$$

where $\cdot$ in the first line denotes multiplication. Note that the probability of the stain being a Caucasian has dropped from a prior probability of 0.819 to a posterior probability of 0.245. This is because of the relative rarity of the profile of X in

the Caucasian population. It can be checked that

$$\Pr(\Psi_2 \mid X) = 0.121, \quad \Pr(\Psi_3 \mid X) = 0.634.$$

Thus, for the Western Polynesian sub-population, the prior probability has increased from 0.044 to 0.634. This is because the profile X is comparatively common in the Western Polynesian sub-population. Note the three probabilities $\Pr(\Psi_1 \mid X) + \Pr(\Psi_2 \mid X) + \Pr(\Psi_3 \mid X) = 1.000$ as they should since the three events $(\Psi_1 \mid X)$, $(\Psi_2 \mid X)$, and $(\Psi_3 \mid X)$ are complementary.

Example 2.5. Another example to illustrate the difference between the two probability statements $\Pr(S \mid R)$ and $\Pr(R \mid S)$ has been provided by Darroch (1987). Consider a town in which a rape has been committed. There are 10 000 men of suitable age in the town of whom 200 work underground at a mine. Evidence is found at the crime scene from which it is determined that the criminal is one of the 200 mineworkers. Such evidence may be traces of minerals that could only have come from the mine. A PoI is identified and traces of minerals, similar to those found at the crime scene are found on some of their clothing. How might this evidence be assessed?

Denote the evidence by E: the event that 'mineral traces have been found on clothing of the PoI which is similar to mineral traces found at the crime scene'. Denote the proposition that the PoI is guilty by H_p and the proposition that they

are innocent by H_d (these are complementary propositions: one and only one is true).

A proposition may be thought of in a similar way to an event, if subjective probabilities are considered. Events may be measurements of characteristics of interest, such as concentrations of certain minerals within the traces. There may be a well-specified model representing the randomness in such measurements. However, the guilt or innocence of the suspect is something about which there is no well-specified model but about which it is perfectly reasonable for an individual to represent with a probability their state of uncertainty about the truth or otherwise of the propositions, as discussed in Section 1.7.5.[3]

Assume that all people working underground at the mine will have mineral traces similar to those found at the crime scene on some of their clothing. This assumption is open to question but the point about conditional probabilities will still be valid. The probability of finding the evidence on an innocent person may then be determined as follows. There are 9999 innocent men in the town of whom 199 work underground at the mine. These 199 men will, as a result of their work, have this evidence on their clothing, under the aforementioned assumption.

[3]Miles (2007) refers to the following Savage's quote to illustrate the link between probability and propositions: 'Personalistic views hold that probability measures the confidence that a particular individual has in the truth of a particular proposition, for example, the proposition that it will rain tomorrow.'

Thus $\Pr(E \mid H_d) = 199/9999 \simeq 200/10\,000 = 0.02$, a small number. Does this imply that a man who is found to have the evidence on him is innocent with probability 0.02? Not at all. There are 200 men in the town with the evidence (E) on them of whom 199 are innocent (H_d). Thus $\Pr(H_d \mid E) = 199/200 = 0.995$. The equation of $\Pr(E \mid H_d)$ with the probability $\Pr(H_d \mid E)$ is known as the *fallacy of the transposed conditional* (Diaconis and Freedman, 1981) and is discussed in more detail later in Sections 2.5.1 and 2.7.1. It has been suggested that this fallacy is about cognitive illusions; see Tversky and Kahneman (1974) and Piattelli-Palmarini (1994) for a general discussion on the phenomenon.

A version of Bayes' theorem for continuous data is introduced in Section 7.3.

2.3 THE ODDS FORM OF BAYES' THEOREM

2.3.1 Likelihood Ratio

Replace S by $\bar{S}$ in (2.4) and the equivalent version of Bayes' theorem is

$$\Pr(\bar{S} \mid R) = \frac{\Pr(R \mid \bar{S}) \Pr(\bar{S})}{\Pr(R)} \qquad (2.6)$$

$(\Pr(R) \neq 0)$.

The first equation (2.4) divided by the second (2.6) gives the *odds form of Bayes' theorem*

$$\frac{\Pr(S \mid R)}{\Pr(\bar{S} \mid R)} = \frac{\Pr(R \mid S)}{\Pr(R \mid \bar{S})} \times \frac{\Pr(S)}{\Pr(\bar{S})}. \qquad (2.7)$$

The left-hand side is the odds in favour of S, given R has occurred. The right-hand side is the product of two terms,

$$\frac{\Pr(R \mid S)}{\Pr(R \mid \bar{S})} \text{ and } \frac{\Pr(S)}{\Pr(\bar{S})}.$$

The latter of these is the odds in favour of S, without any information about R. The former is a ratio of probabilities but it is not in the form of odds. The conditioning events, S and $\bar{S}$, are different in the numerator and denominator, whereas the event R, the probability of which is of interest, is the same.

In the odds form of Bayes' theorem, given here, the odds in favour of S are changed on receipt of information R by multiplication by the ratio $\{\Pr(R \mid S)/\Pr(R \mid \bar{S})\}$. This ratio is important in the evaluation of evidence and is given the name *likelihood ratio* or *Bayes' factor*. Note that a likelihood ratio and a Bayes' factor (BF) are not necessarily equivalent, as in this case where they are taken as synonymous. A discussion about this topic will be provided in Section 2.3.2 and in Section 7.9.

Consider again two events, R and S, and their complements. A likelihood ratio in this context

is the ratio of two probabilities, the probability of R when S is true and the probability of R when S is false. Thus, to consider the effect of R on the odds in favour of S, i.e. to change $\Pr(S)/\Pr(\bar{S})$ to $\Pr(S \mid R)/\Pr(\bar{S} \mid R)$, the former is multiplied by the likelihood ratio. The odds $\Pr(S)/\Pr(\bar{S})$ are known as the *prior odds* in favour of S; i.e. odds prior to receipt of R. The odds $\Pr(S \mid R)/\Pr(\bar{S} \mid R)$ are known as the *posterior odds* in favour of S; i.e. odds posterior to receipt of R[4] With similar terminology, $\Pr(S)$ is known as the *prior probability* of S and $\Pr(S \mid R)$ is known as the *posterior probability* of S. Notice that to calculate the change in the odds on S, it is probabilities of R that are needed. The difference between $\Pr(R \mid S)$ and $\Pr(S \mid R)$, as explained in Section 2.5.1), is vital. Consider two examples.

(1) $\Pr(R \mid S)/\Pr(R \mid \bar{S}) = 3$; the event R is three times more likely if S is true than if S is false. The prior odds in favour of S are multiplied by a factor of 3.

(2) $\Pr(R \mid S)/\Pr(R \mid \bar{S}) = 1/3$; the event R is three times more likely if S is false than if S is true. The prior odds in favour of S are reduced by a factor of 3.

When considering the effect of R on S, it is necessary to consider both the probability of R when S

[4]The word *odds* is sometimes used loosely in reference to the ratio of the probabilities of two mutually exclusive events whose probabilities sum to something different from 1.

is true *and* the probability of R when S is false. It is a frequent mistake (the fallacy of the transposed conditional, Section 2.5.1) to consider that an event R that is unlikely if $\bar{S}$ is true thus provides evidence in favour of S. For this to be so, it is required additionally that R is not so unlikely when S is true. The likelihood ratio is then greater than 1 and the posterior odds are greater than the prior odds.

Notice that the likelihood ratio is a ratio of probabilities. It is greater than zero (except when $\Pr(R \mid S) = 0$ in which case it is zero also) but has no theoretical upper limit. Probabilities take values between 0 and 1, inclusive; the likelihood ratio takes values between 0 and ∞. It is not an odds, however. Odds are the ratio of the probabilities of two complementary events, perhaps conditioned on some other event. The likelihood ratio is the ratio of the probability of the same event conditioned upon two exclusive events, though they need not necessarily be complementary. Thus, $\Pr(R \mid S)/\Pr(R \mid \bar{S})$ is a likelihood ratio; $\Pr(S \mid R)/\Pr(\bar{S} \mid R)$ is an odds statistic.

Kingston and Kirk (1964) developed an example for the likelihood ratio.

> *Now consider a problem of evaluating the significance of several properties in two pieces of glass. Suppose that the probability of two fragments from different sources having this coincidence of properties is .005, and that the probability of such coincidence when they are from the same source is .999. What do these figures mean? They are simply guides for making a decision about the origin of the fragments.*[5] *(p. 514)*

[5]Examples of applications will be found in Chapters 6 and 7.

Notice that (2.7) also holds if S and $\bar{S}$ are propositions rather than events. Propositions may be complementary, such as presence (H_p) or absence (H_d) of a PoI from a crime scene, but need not necessarily be so (see, for example, Section 6.1.4.2 where more than two propositions are compared). In general, the two propositions to be compared will be known as competing propositions. The odds $\Pr(S)/\Pr(\bar{S})$ in such circumstances should be explicitly stated as odds in favour of S, relative to $\bar{S}$. In the special case in which the propositions are mutually exclusive and exhaustive, they are complementary. The odds may then be stated as the odds in favour of S, where the relationship to the complementary proposition is implicit.

2.3.2 Bayes' Factor and Likelihood Ratio

The terms 'likelihood ratio' (LR) and 'Bayes' factor' (BF) have been introduced in Section 2.3.1. In forensic science applications they are often treated as synonymous, though after the seminal work of Lindley (1977c), the former has more common usage in forensic science, both in theory and in practice. In order to clarify the distinction between a likelihood ratio and a Bayes' factor, a formal definition of the Bayes' factor is first provided. The Bayes' factor is the primary element in Bayesian methodology for comparing competing propositions. It is defined as the change produced by new evidence (data) in the odds when going from the prior to the posterior distribution

in favour of one proposition to another. Note that this is not to say that a Bayes' factor depends only upon available data, as will be clarified in Sections 2.3.2.1 and 2.3.2.2. An example will be provided in Section 2.3.2.1 to show that the likelihood ratio is the special case of the Bayes' factor when the competing propositions are parametrised by a single parameter (i.e. a simple hypothesis). There may, however, be cases where composite hypotheses are compared, as in the scenario that will be illustrated in Section 2.3.2.2. In such a case, the Bayes' factor is the ratio of two marginal likelihoods under competing propositions and it appears it no longer depends solely on the data. Note that in what follows, the reference to background information I will be omitted to simplify the notation. A more formal development for continuous parameters will be developed in Section 7.9. Examples dealing with biological findings are presented in Section 6.1.4.2. Several examples of derivations of likelihood ratios are provided in Chapter 7 both for evaluative and discriminative purposes.

2.3.2.1 Derivation of the Bayes' Factor: Simple Versus Simple Propositions

For the purpose of illustration, consider the following simplified scenario taken from Taroni et al. (2014c) involving questioned documents. In general, the propositions involved are of the following form:

H_1: Mr X is the author of the questioned document;

H_2: Mr Y is the author of the questioned document.

Imagine one discrete observation, say, z, of a given handwritten characteristic is available for comparison, where $\Pr(z \mid H_i)$, $i = 1, 2$, is the probability of the observation under the competing propositions. Note that, for ease of notation, $H_1 = \theta_1$ will be taken to mean that Mr X is the author of the questioned document (H_1 holds), whilst $H_2 = \theta_2$ will be taken to mean that Mr Y is the author of the questioned document (H_2 holds).

Following the definition given earlier, the Bayes' factor is the ratio between the posterior odds and the prior odds:

$$\mathrm{BF} = \frac{\Pr(H_1 \mid z)/\Pr(H_2 \mid z)}{\Pr(H_1)/\Pr(H_2)}. \tag{2.8}$$

The posterior probability for H_1 is given by

$$\Pr(H_1 \mid z) = \frac{\Pr(z \mid \theta_1)\Pr(\theta_1)}{\Pr(z \mid \theta_1)\Pr(\theta_1) + \Pr(z \mid \theta_2)\Pr(\theta_2)}.$$

The posterior probability for H_2 is given by

$$\Pr(H_2 \mid z) = \frac{\Pr(z \mid \theta_2)\Pr(\theta_2)}{\Pr(z \mid \theta_1)\Pr(\theta_1) + \Pr(z \mid \theta_2)\Pr(\theta_2)}.$$

Thus, the posterior odds becomes

$$\frac{\Pr(H_1 \mid z)}{\Pr(H_2 \mid z)} = \frac{\Pr(z \mid \theta_1)\Pr(\theta_1)}{\Pr(z \mid \theta_2)\Pr(\theta_2)}.$$

Given that the BF is the ratio between posterior odds to prior odds, one can reduce (2.8) to

$$\begin{aligned}\text{BF} &= \frac{\Pr(H_1 \mid z)}{\Pr(H_2 \mid z)}\frac{\Pr(H_2)}{\Pr(H_1)} \\ &= \frac{\Pr(z \mid \theta_1)}{\Pr(z \mid \theta_2)}\frac{\Pr(\theta_1)}{\Pr(\theta_2)}\frac{\Pr(\theta_2)}{\Pr(\theta_1)} \\ &= \frac{\Pr(z \mid \theta_1)}{\Pr(z \mid \theta_2)},\end{aligned}$$

which is the likelihood ratio.

2.3.2.2 Derivation of the Bayes' Factor: Simple Versus Composite Propositions

Imagine an alternative scenario involving again questioned documents where the competing propositions are as follows:

H_1: Mr X is the author of the questioned document;

H_2: Mr Y or Mr T is the author of the questioned document.

The second proposition, i.e. the defence proposition, may be thought of as a composite proposition; there is more than one possible alternative to the first proposition. Under the latter proposition, H_2, someone else is the author of the document. The term 'someone else' refers to a limited set of authors as the potential source. As in the previous section, $H_1 = \theta_1$ will be taken to mean that Mr X is the author of the questioned document (H_1 holds), whilst $H_2 = \{\theta_2, \theta_3\}$ will be taken to mean that either Mr Y (θ_2) or Mr T (θ_3) is the author of the questioned document (H_2 holds).

Following the same line of reasoning as in the previous section, the posterior probability for H_1 is

$$\Pr(H_1 \mid z) = \frac{\Pr(z \mid \theta_1)\Pr(\theta_1)}{\begin{array}{c}\Pr(z \mid \theta_1)\Pr(\theta_1) + \Pr(z \mid \theta_2)\Pr(\theta_2) \\ + \Pr(z \mid \theta_3)\Pr(\theta_3)\end{array}}.$$

Analogously,

$$\Pr(H_2 \mid z) = \frac{\Pr(z \mid \theta_2)\Pr(\theta_2) + \Pr(z \mid \theta_3)\Pr(\theta_3)}{\begin{array}{c}\Pr(z \mid \theta_1)\Pr(\theta_1) + \Pr(z \mid \theta_2)\Pr(\theta_2) \\ + \Pr(z \mid \theta_3)\Pr(\theta_3)\end{array}}.$$

The posterior odds becomes

$$\frac{\Pr(H_1 \mid z)}{\Pr(H_2 \mid z)} = \frac{\Pr(z \mid \theta_1)\Pr(\theta_1)}{\sum_{i=2,3}\Pr(z \mid \theta_i)\Pr(\theta_i)}.$$

Given that the prior odds are

$$\frac{\Pr(H_1)}{\Pr(H_2)} = \frac{\Pr(\theta_1)}{\sum_{i=2,3}\Pr(\theta_i)},$$

the BF becomes

$$\begin{aligned}\mathrm{BF} &= \frac{\Pr(z \mid \theta_1)\Pr(\theta_1)}{\sum_{i=2,3}\Pr(z \mid \theta_i)\Pr(\theta_i)} \frac{\sum_{i=2,3}\Pr(\theta_i)}{\Pr(\theta_1)} \\ &= \frac{\Pr(z \mid \theta_1)\sum_{i=2,3}\Pr(\theta_i)}{\sum_{i=2,3}\Pr(z \mid \theta_i)\Pr(\theta_i)}.\end{aligned}$$

It can be observed that the Bayes' factor is also a function of the prior inputs and does not simplify to a likelihood ratio. Its interpretation is, however, unchanged.

There is further discussion on this topic in Section 6.1.4 where the composite proposition is split into multiple propositions.

2.3.3 Three-Way Tables

This hypothetical example is based on fingerprint data from thumbs, rounded for ease of understanding, extracted from Champod et al. (2016b). The data are presented in a three-way table with binary characteristics:

- The gender of the person to whom the print belongs: F = female, M = male.
- The hand from which the print was taken: LT = left thumb, RT = right thumb.
- The type of feature: PA = plain arch, TA = tented arch.

Table 2.5 gives the frequency counts extracted from Tables A2 and A3 of Champod et al. (2016b), with the counts edited for ease of arithmetic and reported in tens of thousands.

From Table 2.5, the following probabilities, among others, may be derived.

$$\begin{aligned} \Pr(F) &= \Pr(\text{Female}) = 62/207 = 0.30. \\ \Pr(\text{PA}) &= \Pr(\text{Plain arch}) = 185/207 = 0.89. \end{aligned}$$

$$\Pr(F) \times \Pr(\text{PA}) = \frac{62}{207} \times \frac{185}{207} = 0.268.$$

However,

$$\begin{aligned} \Pr(F \text{ and PA}) &= \Pr(\text{Female and Plain arch}) \\ &= 57/207 = 0.275. \end{aligned}$$

Thus

$$\Pr(F \text{ and PA}) \neq \Pr(F) \times \Pr(\text{PA}).$$

Table 2.5 Frequency of fingerprint characteristics by gender and handedness extracted and edited from Champod et al. (2016b). Reprinted with permissions of CRC Press LLC.

Type of arch	Gender		Total
	Female *F*	Male *M*	
Left thumb (LT)			
Plain arch (PA)	23	45	68
Tented arch (TA)	2	7	9
Total	25	52	77
Right thumb (RT)			
Plain arch (PA)	34	83	117
Tented arch (TA)	3	10	13
Total	37	93	130
Combined			
Plain arch (PA)	57	128	185
Tented arch (TA)	5	17	22
Total	62	145	207

Data are in 10 000s; for example, the entry 22 represents 220 000 prints.

The gender of the print and the characteristic are not independent.

As well as the aforementioned joint probability, conditional probabilities can also be determined. Sixty-two people are female (F). Of these, 57 have a plain arch (PA) Thus, $\Pr(\text{PA} \mid F) \times \Pr(F) =$

$(57/62) \times (62/207) = 57/207 = \Pr(\text{F and PA})$, which provides verification of the third law of probability (1.8). Similarly, $\Pr(F \mid \text{LT}) = 25/77 = 0.325$. Equation (2.4) may be verified by noting that $\Pr(\text{LT}) = 77/207$ and so,

$$\Pr(F \mid \text{LT}) = \frac{\Pr(\text{LT} \mid \text{F}) \times \Pr(F)}{\Pr(\text{LT})} = \frac{(25/62) \times (62/207)}{77/207} = 25/77.$$

Now, consider only people with a plain arch.

$$\begin{aligned} \Pr(F \mid \text{PA}) &= 57/185 \\ &= 0.308. \\ \Pr(\text{LT} \mid \text{PA and } F)) &= 23/(23 + 34) \\ &= 0.404, \end{aligned}$$

and

$$\Pr(F \text{ and LT} \mid \text{PA}) = \Pr(\text{F} \mid \text{PA and LT}) \times \Pr(\text{LT} \mid \text{PA}) = (23/68) \times 68/185 = 23/185$$

from (1.8).

The odds version of Bayes' theorem,

$$\frac{\Pr(F \mid \text{PA})}{\Pr(M \mid \text{PA})} = \frac{\Pr(\text{PA} \mid \text{F}) \times \Pr(\text{F})}{\Pr(\text{PA} \mid \text{M}) \times \Pr(\text{M})} = \frac{\Pr(\text{PA} \mid \text{F})}{\Pr(\text{PA} \mid \text{M})} \times \frac{\Pr(\text{F})}{\Pr(\text{M})},$$

which is of particular relevance to the evaluation of evidence, may also be verified. From Table 2.5 $\Pr(F) = 62/207$, $\Pr(M) = 145/207$. For a print with a plain arch, $\Pr(\text{PA} \mid F) = 57/62$, $\Pr(\text{PA} \mid M) = 128/145$. Thus

$$\begin{aligned}\frac{\Pr(\text{PA} \mid \text{F})}{\Pr(\text{PA} \mid \text{M})} \times \frac{\Pr(F)}{\Pr(M)} &= \frac{(57/62) \times (62/207)}{(128/145) \times (145/207)} \\ &= 57/128 \\ &= \frac{57/185}{128/185} = \frac{\Pr(F \mid \text{PA})}{\Pr(M \mid \text{PA})}.\end{aligned} \tag{2.9}$$

The odds form of Bayes' theorem has been verified numerically.

Consider this example as an identification problem. Before any information about a fingermark, say, is available, the pool of people of interest is known to contain 62 females and 145 males. The odds in favour of the criminal being female are 62:145 (62/145). It is then discovered that there is a plain arch in a fingermark determined to have come from the criminal. The odds are now changed, using the procedure in the example, from being 62 : 145 in favour of female to 57 : 128 in favour of female. The effect or value of the evidence has been to multiply the prior odds by a factor of 1.04. The value of the evidence is 1.04. The probability the criminal is female

is $57/185 = 0.308$. The prior probability the criminal was female was $62/207 = 0.300$.

2.3.4 Logarithm of the Likelihood Ratio

Odds and the likelihood ratio take values between 0 and ∞. Logarithms of these statistics take values on $(-\infty, \infty)$. Also, the odds form of Bayes' theorem involves a multiplicative relationship. If logarithms are taken, the relationship becomes an additive one:

$$\log\left\{\frac{\Pr(S \mid R)}{\Pr(\bar{S} \mid R)}\right\} = \log\left\{\frac{\Pr(R \mid S)}{\Pr(R \mid \bar{S})}\right\} + \log\left\{\frac{\Pr(S)}{\Pr(\bar{S})}\right\}. \quad (2.10)$$

The idea of evaluating evidence by adding it to the logarithm of the prior odds is very much in keeping with the intuitive idea of weighing evidence in the scales of justice. The logarithm of the likelihood ratio has been given the name *the weight of evidence* (Good, 1950), and see also Peirce (1878). Peirce defines *probability* as the ratio of 'favourable cases to all the cases'. He defined the *chance* of an event the ratio of 'favourable cases to unfavourable cases', what is now known as odds. He argued that the chance of an event has 'an intimate connection with our belief in it'. Any

'quantity which varies with the chance might, therefore, it would seem serve as a thermometer for the proper intensity of belief'. He then argued that the quantity that serves as a thermometer in this way is the 'logarithm of the chance'. He mentioned another consideration to support the choice of a logarithm. 'It is that our belief ought to be proportional to the weight of evidence'. Finally, Peirce (1878) wrote:

> *The rule for the combination of independent concurrent arguments takes a very simple form when expressed in terms of the intensity of belief, measured in the proposed way [with logarithms]. It is this: Take the sum of all the feelings of belief which would be produced separately by all the arguments* pro, *subtract from that the similar sum for arguments* con *and the remainder is the feeling of belief we ought to have on the whole. (p. 294)*

A likelihood ratio with a value greater than 1, which leads to an increase in the odds in favour of *S*, has a positive weight. A likelihood ratio with a value less than 1, which leads to a decrease in the odds in favour of *S*, has a negative weight. A positive weight may be thought to tip the scales of justice one way, a negative weight may be thought to tip the scales of justice the other way. A likelihood ratio with a value equal to 1 leaves the odds in favour of *S* and the scales unchanged. The evidence is logically relevant only when the probability of finding that evidence given the truth of some proposition at issue in the case differs

from the probability of finding the same evidence given the falsity of the proposition at issue; i.e. the log-likelihood ratio is not zero (Kaye, 1986b). The logarithm of the likelihood ratio (sometimes called *relevance ratio* or *weight*) provides an equivalent measure of relevance. This method is advantageous, because it equates the relevance of evidence offered by both the prosecution and the defendant (Lempert, 1977). The log-likelihood ratio also has qualities of symmetry and additivity that other measures lack (Edwards, 1986). Lyon and Koehler (1996) believe the simplicity and intuitive appeal of the relevance ratio make it a good candidate for heuristic use by judges. The mathematical symmetry between the weight of evidence for the prosecution proposition and the weight of evidence for the defendant's proposition can be maintained by inverting the weight of evidence when considering the defendant's proposition.

A verbal scale based on logarithms (Aitken and Taroni, 1998) and several other verbal scales are discussed in Section 2.4.6.

Example 2.6. Consider two propositions regarding a coin.

- S: the coin is double-headed; if S is true, the probability of tossing a head equals 1, the probability of tossing a tail equals 0.
- $\bar{S}$: the coin has one head and one tail and is fair: if $\bar{S}$ is true the probability of tossing a head equals

the probability of tossing a tail and both equal 1/2.

Notice that these are not complementary propositions; the coin may be biased.

The coin is tossed 10 times and the outcome of any one toss is assumed independent of the others. The result R is 10 heads. Then $\Pr(R \mid S) = 1$, $\Pr(R \mid \bar{S}) = (\frac{1}{2})^{10}$. The likelihood ratio is

$$\frac{\Pr(R \mid S)}{\Pr(\bar{R} \mid \bar{S})} = \frac{1}{(1/2)^{10}} = 2^{10} = 1024.$$

The evidence is 1024 times more likely if the coin is double-headed. The weight of the evidence is 10 log(2). Each toss that yields a head contributes a weight log(2) to the hypothesis S that the coin is double-headed. Suppose, however, that the outcome of one toss is a tail (T). Then $\Pr(T \mid S) = 0$, $\Pr(T \mid \bar{S}) = 1/2$. The likelihood ratio $\Pr(T \mid S)/\Pr(T \mid \bar{S}) = 0$ and the posterior odds in favour of S relative to $\bar{S}$ equals 0. This is to be expected. A double-headed coin cannot produce a tail. If a tail is the outcome of a toss then the coin cannot be double-headed.

A brief history of the use of the weight of evidence is given by Good (1991). Units of measurement are associated with it. When the base of the logarithms is 10, Turing suggested that the unit should be called a *ban* and one tenth of this would be called a *deciban*, abbreviated to *db*, and thought equivalent to the minimum value of evidence which it is possible to detect in an investigation.

2.4 THE VALUE OF EVIDENCE

2.4.1 Evaluation of Forensic Evidence

Consider the odds form of Bayes' theorem in the forensic context of assessing the value of some evidence. The initial discussion is in the context of the guilt or otherwise of the PoI. This may be the case, for example, in the context of Example 1.1, if all innocent explanations for the bloodstain have been eliminated. Later, greater emphasis will be placed on propositions that the PoI was, or was not, present at the scene of the crime. At present, replace event S by a proposition H_p, that the PoI (or defendant if the case has come to trial) is truly guilty. Event $\bar{S}$ is replaced by proposition H_d, that the PoI is truly innocent. Event R is replaced by event Ev, the evidence under consideration. This may be written as $(E, M) = (E_c, E_s, M_c, M_s)$, the type of evidence and observations of it as described in Section 1.7.1. The odds form of Bayes' theorem then enables the prior odds (i.e. prior to the presentation of Ev) in favour of guilt to be updated to posterior odds given Ev, the evidence under consideration. This is done by multiplying the prior odds by the likelihood ratio that, in this context, is the ratio of the probabilities of the evidence assuming guilt and assuming innocence of the PoI. With this notation, the odds form of Bayes' theorem may be written as

$$\frac{\Pr(H_p \mid Ev)}{\Pr(H_d \mid Ev)} = \frac{\Pr(Ev \mid H_p)}{\Pr(Ev \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}.$$

Explicit mention of the background information I is omitted in general from probability statements for ease of notation. With the inclusion of I the odds form of Bayes' theorem is

$$\frac{\Pr(H_p \mid Ev, I)}{\Pr(H_d \mid Ev, I)} = \frac{\Pr(Ev \mid H_p, I)}{\Pr(Ev \mid H_d, I)} \times \frac{\Pr(H_p \mid I)}{\Pr(H_d \mid I)}.$$

Notice the important point that in the evaluation of the evidence Ev it is two probabilities that are necessary: the probability of the evidence if the PoI is guilty and given the background information and the probability of the evidence if the PoI is innocent and given the background information. Background information is sometimes known as the *framework of circumstances* or the *conditioning information*. As noticed in the ENFSI Guideline for evaluative reporting in forensic science (ENFSI, 2015), the background information is

> *[...] the relevant case information that helps the forensic practitioner recognise the pertinent issues, select the appropriate propositions and carry out the case pre-assessment.*[6] *It shall always be regarded as provisional and the examiner shall be ready to re-evaluate findings if the conditioning information changes. Examples of relevant information that could change include the nature of the alleged activities, time interval between incident and the collection of traces (and reference items) and the suspect's/victim's account of their activities. (p. 21)*

More details on the role of background information are presented in Section 2.4.4.

[6]See Sections 5.5.1 and 5.5.2 for details on case and evidence pre-assessment, respectively.

Note that it is not sufficient to consider only the probability of the evidence if the PoI is innocent and to declare that a small value of this is indicative of guilt. The probability of the evidence if the PoI is guilty has also to be considered.

Similarly, it is not sufficient to consider only the probability of the evidence if the PoI is guilty and to declare that a high value of this is indicative of guilt. The probability of the evidence if the PoI is innocent has also to be considered. An example of this is the treatment of the evidence of a bite mark in the Biggar murder in 1967–1968 (Harvey et al., 1968), an early example of odontology in forensic science. In that murder a bite mark was found on the breast of the victim, a young girl, which had certain characteristic marks, indicative of the conformation of the teeth of the person who had bitten her. A 17-year old boy, who was already a PoI, was found with this conformation. Such evidence would help towards the calculation of $\Pr(Ev \mid H_p, I)$. However, there was no information available about the incidence of this conformation among the general public. Examination was made of 90 boys aged 16–18. This enabled an estimate – albeit an intuitive one – of $\Pr(Ev \mid H_d, I)$ to be obtained and to show that the particular conformation found on the breast of the victim was not at all common.

Consider the likelihood ratio $\Pr(Ev \mid H_p)/ \Pr(Ev \mid H_d)$ further where explicit mention of I has again

been omitted. This equals

$$\frac{\Pr(E \mid H_p, M)}{\Pr(E \mid H_d, M)} \times \frac{\Pr(M \mid H_p)}{\Pr(M \mid H_d)}.$$

The second ratio in this expression, $\Pr(M \mid H_p)/\Pr(M \mid H_d)$, concerns the type and quantity of evidential material found at the crime scene and on the suspect. It may be written as

$$\frac{\Pr(M_c \mid M_r, H_p)}{\Pr(M_c \mid M_r, H_d)} \times \frac{\Pr(M_r \mid H_p)}{\Pr(M_r \mid H_d)}.$$

The value of the second ratio in this expression may be taken to be 1. The type and quantity of material at the crime scene is independent of whether the PoI is the criminal or someone else is. The value of the first ratio, which concerns the evidential material found on the PoI given the evidential material found at the crime scene and the guilt or otherwise of the PoI, is a matter for subjective judgement and it is not proposed to consider its determination further here. Instead, consideration will be concentrated on

$$\frac{\Pr(E \mid H_p, M)}{\Pr(E \mid H_d, M)}.$$

In particular, M will be subsumed into I and omitted, for ease of notation. Then,

$$\frac{\Pr(M \mid H_p)}{\Pr(M \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

which equals

$$\frac{\Pr(H_p \mid M)}{\Pr(H_d \mid M)}$$

will be written as

$$\frac{\Pr(H_p)}{\Pr(H_d)}.$$

Thus

$$\frac{\Pr(H_p \mid Ev)}{\Pr(H_d \mid Ev)} = \frac{\Pr(H_p \mid E, M)}{\Pr(H_d \mid E, M)}$$

will be written as

$$\frac{\Pr(H_p \mid E)}{\Pr(H_d \mid E)}$$

and

$$\frac{\Pr(Ev \mid H_p)}{\Pr(Ev \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

will be written as

$$\frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}.$$

The full result is then

$$\frac{\Pr(H_p \mid E)}{\Pr(H_d \mid E)} = \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}, \qquad (2.11)$$

or if I is included

$$\frac{\Pr(H_p \mid E, I)}{\Pr(H_d \mid E, I)} = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)} \times \frac{\Pr(H_p \mid I)}{\Pr(H_d \mid I)}. \qquad (2.12)$$

The likelihood ratio is the ratio

$$\frac{\Pr(H_p \mid E, I)/\Pr(H_d \mid E, I)}{\Pr(H_p \mid I)/\Pr(H_d \mid I)} \tag{2.13}$$

of posterior odds to prior odds. It is the factor that converts the prior odds in favour of guilt to the posterior odds in favour of guilt or more generally, it is the factor that converts the prior odds in favour of a proposition put forward by a given party, say, the prosecutor, to the posterior odds in favour of the same proposition. The representation in (2.12) also emphasises the dependence of the prior odds on background information. Previous evidence may be included here also, see, for example, Section 5.3.2.5.

It is often the case that another representation may be appropriate. Sometimes it may not be possible to consider the effect of the evidence on the guilt or innocence of the PoI. However, it may be possible to consider the effect of the evidence on the possibility that a given action has been committed by the PoI at the crime scene. For example, a blood stain at the crime scene may have the same DNA profile as that of the PoI. This, considered in isolation, would not necessarily be evidence to suggest that the PoI was guilty, only that they were at the crime scene. Consider the following two complementary propositions:

H_p: the PoI was at the crime scene;

H_d: the PoI was not at the crime scene.

The odds form of Bayes' theorem is then

$$\frac{\Pr(H_p \mid E, I)}{\Pr(H_d \mid E, I)} = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)} \times \frac{\Pr(H_p \mid I)}{\Pr(H_d \mid I)}, \quad (2.14)$$

identical to (2.12) but with different definitions for H_p and H_d. The likelihood ratio converts the prior odds in favour of H_p into the posterior odds in favour of H_p.

The likelihood ratio may be thought of as the *value* of the evidence. Evaluation of evidence, the theme of this book, will be taken to mean the determination of a value for the likelihood ratio. This value will be denoted V.

Definition. Consider two competing propositions, H_p and H_d, and background information I. The *value* V of the evidence E is given by

$$V = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}, \quad (2.15)$$

the likelihood ratio that converts prior odds $\Pr(H_p \mid I)/\Pr(H_d \mid I)$ in favour of H_p relative to H_d into posterior odds $\Pr(H_p \mid E, I)/\Pr(H_d \mid E, I)$ in favour of H_p relative to H_d.

An illustration of the effect of evidence with a value V of 1000 on the odds in favour of H_p, relative to H_d, is given in Table 2.6.

This is not a new idea. Consider the following quotes from Kaye (1979):

Table 2.6 Effect on prior odds in favour of H_p relative to H_d of evidence E with value V of 1 000.

Prior odds $\Pr(H_p)/\Pr(H_d)$	V	Posterior odds $\Pr(H_p \mid E)/\Pr(H_d \mid E)$
1/10 000	1 000	1/10
1/100	1 000	10
1 (evens)	1 000	1 000
100	1 000	100 000

Reference to background information I is omitted.

> *That approach does not ask the jurors to produce any number, let alone one that can qualify as a probability. It merely shows them how a 'true' prior probability would be altered, if one were in fact available. It thus supplies the jurors with as precise and accurate an illustration of the probative force of the quantitative data as the mathematical theory of probability can provide. Such a chart, it can be maintained, should have pedagogical value for the juror who evaluates the entire package of evidence solely by intuitive methods, and who does not himself attempt to assign a probability to the 'soft' evidence.*
>
> *[A] more fundamental response is that there appears to be no reason in principle why a juror could not generate a prior probability that could be described in terms of the objective, relative-frequency sort of probability. One could characterise the juror 's prior probability as an estimate of the proportion of cases in which a defendant confronted with the same pattern of non-quantitative evidence as exists in the case at bar would in fact turn out to have stabbed the deceased.*
>
> *This practical difficulty does not undercut the conceptual point. (pp. 52–53)*

This comment was reiterated by Kaye (1982) by affirming:

> *[one] could merely display posterior distributions for a wide range of possible priors, for the purpose of showing the probative force of the measured value of X. This procedure would not require the court or jury to settle on any particular prior distribution. (p. 779)*

This idea was supported by Fienberg (1982) followed by Berry and Geisser (1986).

Care has to be taken with the interpretation of the likelihood ratio if the propositions H_p and H_d are not exhaustive. Consider the following example taken from Royall (1997, p. 8). Evidence E that supports H_p over H_d in the sense that the likelihood ratio $\Pr(E \mid H_p)/\Pr(E \mid H_d)$ is greater than 1 can be such that the probabilities of both propositions are reduced; $\Pr(H_p \mid E) < \Pr(H_p)$ and $\Pr(H_d \mid E) < \Pr(H_d)$. Suppose there is a third proposition H_c and that *a priori* $\Pr(H_p) = \Pr(H_d) = \Pr(H_c) = \frac{1}{3}$ so that the three propositions are mutually exclusive and exhaustive; propositions H_p and H_d are exclusive but not exhaustive. Suppose also that there is evidence E that takes a value x, such that $\Pr(H_p \mid x) = \frac{1}{6}, \Pr(H_d \mid x) = \frac{1}{12}$ and $\Pr(H_c \mid x) = \frac{3}{4}$. The effect of E is reduce the probability of H_p and of H_d whilst increasing the probability of H_c. Evidence E also doubles the probability of

H_p relative to H_d; i.e. $\Pr(H_p \mid E = x) < \Pr(H_p)$ and $\Pr(H_d \mid E = x) < \Pr(H_d)$, yet

$$\frac{\Pr(H_p \mid E = x)}{\Pr(H_d \mid E = x)} = 2\,\frac{\Pr(H_p)}{\Pr(H_d)}.$$

The evidence does not support H_p taken alone – it supports H_p over H_d. This caveat should be made clear in the interpretation of evidence.

Similar arguments hold for other values of the likelihood ratio when the two propositions are not exhaustive. In particular, it is possible to have evidence E with a likelihood ratio of 1 and $\Pr(H_p \mid E) > \Pr(H_p)$. This inequality does not mean that E is probative for H_p since $\Pr(H_d \mid E) > \Pr(H_d)$ also and

$$\frac{\Pr(H_p \mid E)}{\Pr(H_d \mid E)} = \frac{\Pr(H_p)}{\Pr(H_d)}.$$

For further discussion see Fenton et al. (2014) and Biedermann et al. (2014).

A further property of the likelihood ratio is the ease with which the posterior odds may be updated with new evidence. The posterior odds for one piece of evidence become the prior odds for the next piece of evidence. Given two pieces of evidence, E_1 and E_2,

$$\frac{\Pr(H_p \mid E_1, E_2)}{\Pr(H_d \mid E_1, E_2)} = \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} \times \frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)}. \tag{2.16}$$

An application of the above argument can be illustrated from *R v. Adams, D.J.* (1997) and Dawid (2002).

Adams was arrested for rape. The evidence E linking him to the crime was, first, a match between his DNA and that of semen obtained from the victim and, second, the fact that he lived locally. A conditional match probability of 1 in 200 million for the DNA was reported (see Section 6.1.5 for a comment on such a statistic). The defence challenged this and suggested a figure of 1 in 20 million or even 1 in 2 million could be more appropriate. There was other (prior) information:

- *Identification* (I_1): The victim gave a description of her attacker, which was hard to reconcile with the defendant and did not pick out the defendant in an identity parade.
- *Alibi* (I_2): A former girlfriend of Adams gave an alibi, which was not challenged.

At the trial, with the consent of the prosecution, the defence and the Court, the jury was given instruction in the correct way to combine all the evidence. A prior probability of guilt was introduced, followed by what were claimed to be plausible likelihood ratios for I_1 and I_2. The DNA evidence (E) was then introduced to provide a final posterior probability of guilt. Of course, the figures which follow could be challenged or the jury could substitute their own figures.

Consider two propositions:

H_p: Adams is guilty;

H_d: Adams is not guilty.

The prior probability of guilt was assessed as follows: there were thought to be approximately 150 000 males between 18 and 60 in the area who, in the absence of other evidence, may have committed the crime. Another 50 000 were added to this to allow for the possibility that the attacker may have come from outside the area (i.e. a probability of 0.25 for the possibility that the attacker came from outside the area was thought appropriate). Thus $\Pr(H_p) = 1/200\,000$, $\Pr(H_p)/\Pr(H_d) = 1/199\,999 \simeq 1/200\,000$. The other two pieces of evidence, that of identification (I_1) and alibi (I_2), were assessed as follows:

- *Identification evidence*: I_1 was assigned a probability $\Pr(I_1 \mid H_p) = 0.1$, if he were guilty and a probability $\Pr(I_1 \mid H_d) = 0.9$ if he were innocent. Note that these two probabilities sum to 1 but they need not necessarily do so (see the alibi evidence in the following text.) These assignations provided a likelihood ratio $\Pr(I_1 \mid H_p)/\Pr(I_1 \mid H_d) = 1/9$ (or a likelihood ratio $\Pr(I_1 \mid H_d)/\Pr(I_1 \mid H_p) = 9$ in favour of the defence).
- *Alibi evidence*: I_2 was assigned a probability $\Pr(I_2 \mid H_p) = 0.25$, if he were guilty and

a probability $\Pr(I_2 \mid H_d) = 0.50$, if he were innocent. These assignations provided a likelihood ratio $\Pr(I_2 \mid H_p)/\Pr(I_2 \mid H_d) = 1/2$ (or a likelihood ratio $\Pr(I_2 \mid H_d)/\Pr(I_2 \mid H_p) = 2$ in favour of the defence).

These two items of evidence were assumed to be independent. Thus, placing H_d in the numerator and H_p in the denominator,

$$\begin{aligned}\frac{\Pr(I_1, I_2 \mid H_d)}{\Pr(I_1, I_2 \mid H_p)} &= \frac{\Pr(I_1 \mid H_d)}{\Pr(I_1 \mid H_p)} \times \frac{\Pr(I_2 \mid H_d)}{\Pr(I_2 \mid H_p)} \\ &= 18\end{aligned}$$

which is an overall likelihood ratio of 18 in favour of the defence. This likelihood ratio can then be combined with the prior odds $\Pr(H_d)/\Pr(H_p)$ of 200 000, to give odds before consideration of the DNA evidence of

$$\begin{aligned}\frac{\Pr(H_d \mid I_1, I_2)}{\Pr(H_p \mid I_1, I_2)} &= \frac{\Pr(I_1 \mid H_d)}{\Pr(I_1 \mid H_p)} \times \frac{\Pr(I_2 \mid H_d)}{\Pr(I_2 \mid H_p)} \\ &\quad \times \frac{\Pr(H_d)}{\Pr(H_p)}\end{aligned}$$

which equals 3.6 million in favour of the defence or against the prosecution.

Now the DNA evidence E (with a likelihood ratio of 200 million to 2 million in favour of the prosecution) is included through multiplication with the odds of 1 in 3.6 million, from consideration of I_1 and I_2. The DNA evidence E is assumed

independent of I_1 and I_2.

$$
\begin{aligned}
V &= \frac{\Pr(H_p \mid E, I_1, I_2)}{\Pr(H_d \mid E, I_1, I_2)} \\
&= \frac{\Pr(E, I_1, I_2 \mid H_p)}{\Pr(E, I_1, I_2 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)} \\
&= \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} \times \frac{\Pr(I_1 \mid H_p)}{\Pr(I_1 \mid H_d)} \times \frac{\Pr(I_2 \mid H_p)}{\Pr(I_2 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)},
\end{aligned}
$$

where $\Pr(E \mid H_p)/\Pr(E \mid H_d)$ may be taken to 200 million (the prosecution's suggestion) or 2 million (the lower of the suggestions of the defence). These results give posterior odds from 56 to 1 (200/3.6) in favour of guilt to 1.8 to 1 in favour of innocence. These odds in turn give a posterior probability of guilt of 0.98 (56/57) or 0.36 (1 − 1.8/2.8) depending on the conditional match probability previously considered. The defence argued from these figures that guilt was not proved 'beyond reasonable doubt'.

The jury returned a verdict of guilty. The Appeal Court rejected the attempt to introduce probabilistic reasoning into court saying that 'it trespasses on an area peculiarly and exclusively within the province of the jury', and 'to introduce Bayes' theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task'. (This is reminiscent of comments made in the appeal for the Collins' case, Section 3.4.) The task of the jury was said to be to 'evaluate evidence and

reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them'. This fails to recognise that so-called common sense usually fares very badly when it comes to manipulating probabilities.

The appeal was granted on the basis that the trial judge had not adequately dealt with the question of what the jury should do if they did not want to use Bayes' theorem. A retrial was ordered. Attempts were made to describe the Bayesian approach to the integration of all the evidence. Once again the jury convicted, once again the case went to appeal and once again the Bayesian approach was rejected as inappropriate to the courtroom, and the appeal was dismissed.

Note that Dennis Adams had a full brother whose DNA was not investigated. The probability that the brother had the same DNA profile as the defendant was calculated as 1 in 220. This was claimed to weaken the impact of the DNA evidence against Dennis Adams. This point was dismissed on the grounds that there was no evidence as to the brother's actual DNA profile, nor any suggestion that he might have committed the offence. However, there was also no other evidence against Dennis Adams except the DNA evidence and that he lived locally.

Consider the case of *R v. Lashley* (2000) discussed by Redmayne (2002). Lashley was convicted of the robbery of a Liverpool post office. The only evidence against him was a DNA match,

which left him as a suspect along with 7–10 other males in the United Kingdom. There was no evidence linking him to the Liverpool area. His conviction was quashed on appeal.

However, the case of *R v. Smith* (2000), with the same judges on the same day as *Lashley* (2000), was treated differently. He was convicted of the robbery of a post office in Barley, Hertfordshire. The principal evidence against him was a corresponding DNA profile (with conditional match probability 1 in 1.25 million), which (as in *R v. Adams, D.J.* (1997)) left him as a suspect along with 43 other males in the United Kingdom. His appeal was quashed because the DNA evidence 'did not stand alone because there was also quite clearly evidence of this man having been arrested some shortish distance away'.

Smith was arrested at a place called Potton, which is 13 miles from Barley. A circle centred on Barley with Potton on its border encloses several cities and, at a rough estimate, some 80 000 men of appropriate age who live at least as close to Barley as Smith. This figure could be used to provide prior odds of 1 in 80 000 against Smith's guilt. This may be combined with the DNA evidence to produce a probability of guilt of 0.94. If it is thought at least 16 times as bad to convict an innocent person as to acquit a guilty person, then Smith would not be convicted.

Also, Smith came from a large family – his father had 13 brothers and sisters. The Court of Appeal did not pursue this. As in *R v. Adams, D.J.* (1997),

there was no evidence to implicate the relatives, but, apart from the DNA (and geography) neither was there evidence to implicate Smith. If it is assumed that all members of the population of potential suspects are equally likely to have committed the crime, then any relatives among the population are, prior to the DNA test, as likely to be guilty as anyone else. Relatives are far more likely to match than other members of the population and so should not be ignored (Lempert, 1991, 1993; Balding, 2000; Redmayne, 2002). See also Section 6.1.4 for a discussion on relatives and multiple propositions.

2.4.2 Justification of the Use of the Likelihood Ratio

The odds form of Bayes' theorem presents a compelling intuitive argument for the use of the likelihood ratio as a measure of the value of the evidence, or – as expressed by Good (1985) – a measure to quantify 'the factor in favour of H provided by E' (p. 252).

A mathematical argument exists also to justify its use. A simple proof is given in Good (1989), reiterated in Good (1991) and repeated here for convenience.[7]

It is desired to measure the value V of evidence E in favour of guilt H_p. There will be dependence on background information I but this will not be stated explicitly. It is assumed that this value V is a

[7]Note that a theorem showing that the likelihood ratio measure can be axiomatised has been presented by Crupi et al. (2013).

function only of the probability of E given that the PoI is guilty, H_p, and of the probability of E given that the PoI is innocent, H_d.

Let $x = \Pr(E \mid H_p)$ and $y = \Pr(E \mid H_d)$. The assumption above states that

$$V = f(x, y)$$

for some function f.

Consider another piece of evidence T, which is irrelevant to (independent of) E and H_p (and hence H_d) and which is such that $\Pr(T) = \theta$. Then

$$\Pr(E, T \mid H_p) = \Pr(E \mid H_p) \Pr(T \mid H_p)$$

by the independence of E and T

$$= \Pr(E \mid H_p) \Pr(T)$$

by the independence of T and H_p

$$= \theta x.$$

Similarly,

$$\Pr(E, T \mid H_d) = \theta y.$$

The value of the combined evidence (E, T) is equal to the value of E, since T has been assumed irrelevant. The value of (E, T) is $f(\theta x, \theta y)$ and the value of $E = V = f(x, y)$. Thus

$$f(\theta x, \theta y) = f(x, y)$$

for all θ in the interval $[0,1]$ of possible values of $\Pr(T)$.

The relationship between x and y within the function f may take one of four forms depending on the four mathematical operators $+, \times, -$ and $/$.

+ $f(x, y) = f(x + y)$;
$f(\theta x, \theta y) = f(\theta x + \theta y) = f(\theta(x + y))$. This is not equal to $f(x + y)$ for all θ in the interval $[0, 1]$. For example, try

$$\begin{aligned} f(x, y) &= (x + y)^2; \\ f(\theta x, \theta y) &= \theta^2 (x + y)^2 \\ &= \theta^2 f(x, y). \end{aligned}$$

× $f(x, y) = f(x \times y)$;
$f(\theta x, \theta y) = f(\theta x \times \theta y) = f(\theta^2 (x \times y))$. This is not equal to $f(x \times y)$ for all θ in the interval $[0,1]$.

$$\begin{aligned} f(x, y) &= (x \times y)^2; \\ f(\theta x, \theta y) &= \theta^2 (x \times y)^2 \\ &= \theta^2 f(x, y). \end{aligned}$$

− $f(x, y) = f(x - y)$;
$f(\theta x, \theta y) = f(\theta x - \theta y) = f(\theta(x - y))$. This is not equal to $f(x - y)$ for all θ in the interval $[0,1]$.

$$\begin{aligned} f(x, y) &= (x - y)^2; \\ f(\theta x, \theta y) &= \theta^2 (x - y)^2 \\ &= \theta^2 f(x, y). \end{aligned}$$

/ $f(x, y) = f(x/y)$;
$f(\theta x, \theta y) = f(\theta x/\theta y) = f(x/y)$. This is equal to $f(x, y)$ for all θ in the interval $[0,1]$.

$$\begin{aligned} f(x, y) &= (x/y)^2; \\ f(\theta x, \theta, y) &= (\theta x/\theta y)^2 \\ &= f(x, y). \end{aligned}$$

It follows that f is a function of x/y alone and hence that V is a function of

$$\Pr(E \mid H_p)/\Pr(E \mid H_d),$$

namely, the likelihood ratio.

Arguments for the use of the likelihood ratio as a measure for the value of evidence can be found in Evett (1998), Champod et al. (2016a), Gittelson et al. (2018), Ostrum (2019) and in the Guidelines principles developed under the programme 'Probability and Statistics in Forensic Science' (2017).[8] Recently, the UK Forensic Science Regulator's recommendations regarding image comparison evidence clearly expressed the way to proceed, even for digital evidence (Tully and Stockdale, 2019):

> *What is the probability that those observations would have been made if the first proposition were true? What is the probability that those observations would have been made if the second proposition were true? It is the ratio of those two probabilities that is central to the evaluation. (p. 4)*

It is not unusual for likelihood ratios to be presented by experts in US courts; examples are *People v. Collins* (2015), *People v. Belle* (2015), *People v. Carter* (2016), *People v. Bullard-Daniel* (2016), *Ghebrezghi v. State* (2018), *Commonwealth v. McClellan* (2018), and *United States v. Williams* (2019).

[8] Report available at: www.newton.ac.uk/files/preprints/ni16061.pdf.

2.4.3 Single Value for the Likelihood Ratio

The use of the likelihood ratio or Bayes' factor as a metric to assess the probative value of forensic findings is largely supported by operational standards and recommendations in different forensic disciplines, see, for example, the standards presented by the Association of Forensic Science Providers (2009) and the ENFSI (2015) guideline for evaluative reporting. It must be recognised that the assessment of a value for the likelihood ratio can be subjected to many sources of uncertainty, including the quality of the data obtained from analyses conducted by forensic scientists, the choice of control sample and recovered items that may be taken by different investigators or analysed by different analysts or laboratories. The evaluation of scientific evidence in court often requires a combination of data on the occurrence of target features, together with a personal knowledge of circumstances from a particular case. Clearly, any probability judgement referring to a particular case, even when thought of in frequency format, has a component based on personal knowledge (Taroni et al., 2016, 2018b). Other sources of uncertainty include the elicitation of prior probabilities conditional on available knowledge (see Section 2.3.2 where the possible influence of prior inputs on the value of the evidence is mentioned), or even the implementation of numerical

procedures to overcome computational difficulties (Section 7.6.2.4).

There has been some debate as to whether a likelihood ratio value should be accompanied by an interval to take into account the various sources of uncertainty. There is a school of thought that recommends that a report on the value of the likelihood ratio should include a measure of its precision, for example, through the provision of a numeric range of values for what it is believed to be the probability of the evidence under the competing propositions and therefore a numerical range of values for the likelihood ratio. However, the value of the evidence and the strength of one's belief in the value are different concepts and should not be combined in an interval or result in a change in the value of the evidence, such as is done, for example, with the provision of a lower bound of some arbitrarily chosen level. In practice, for a criminal investigation, there is one set of background data characterising members of a given relevant population, one set of control data, and one set of recovered data. Therefore, for the evaluation of evidence with a particular statistical model, there is one single value V for the associated likelihood ratio. Again, it is to be hoped that any different control sample and recovered items are sufficiently representative of the populations from which they have been selected so that the likelihood ratios will differ little in value. If asked about the trustworthiness

of their reported values, a scientist can answer by describing the amount of data and the available expert's knowledge used in the evaluative process.

An appreciation of the robustness of particular assigned probabilities and, by extension, likelihood ratios, is affected by the quality of the data and the model on which the analysis was conducted. Such explanations should help illustrate the extent to which a particular reported value is data-based. In this way, all the knowledge that informed the stated value of the evidence is made available to the court. It can then be explained that the reported value remains the best quantification possible given (i) the available evidence, (ii) the chosen model, and (iii) the background data. The probability assignments and, consequently, the value of a likelihood ratio, determines the best representation of one's state of uncertainty at a given moment in time. This aspect should not be confused with the quantification of misleading evidence, that is the assessment of the discriminative capacity of a particular examination and evaluation procedure with respect to a particular proposition of interest. This argument about misleading evidence, which is extensively treated in Chapter 8, should however not distract from the given case assessment. See also Section 7.6.3.

A report of a single value does not mean there is no uncertainty, it simply means there is no uncertainty about how to deal with the evidence. If more evidence or data were available, and a new model adopted, then a new analysis may be

conducted which may result in a change in the value. However, these different values should not be combined in an interval. Moreover, it is unclear how the values of different pieces of evidence that are in the form of intervals could be combined in a meaningful way (Bozza et al., 2018).

An extended argument in support of the use a single value likelihood ratio can be found in Taroni et al. (2016). Other scientists agreed with this point of view or reiterated it, see for example, Ommen et al. (2016), Berger and Slooten (2016), Biedermann et al. (2016b), Taylor et al. (2016c), and Dawid (2017), sometimes by addressing some practical difficulties (Nordgaard, 2016). Criticisms have been raised and addressed by Sjerps et al. (2016), Morrison and Enzinger (2016), Curran (2016), and van den Hout and Alberink (2016).

2.4.4 Role of Background Information

From (2.15) it can be seen that the value of the evidence depends on the background information I as well as the propositions H_p and H_d. The background information is personal to each individual. It may be thought, therefore, that the value of the evidence is different for each individual, the posterior odds are then different for each individual and hence there is little point in evaluating evidence. That this is not the case when one individual is a forensic scientist and another is a fact-finder was pointed out by Aitken and Nordgaard (2017).

Denote two participants in the judicial process as A and B and denote the background information available to each as I_a and I_b, respectively, with the overall background information available to them as $I = I_a \cup I_b$. The odds form of Bayes theorem (2.14) may then be written as

$$\frac{\Pr(H_p \mid E, I_a \cup I_b)}{\Pr(H_d \mid E, I_a \cup I_b)} = \frac{\Pr(E \mid H_p, I_a \cup I_b)}{\Pr(E \mid H_d, I_a \cup I_b)} \times \frac{\Pr(H_p \mid I_a \cup I_b)}{\Pr(H_d \mid I_a \cup I_b)}.$$

Partition $I_a \cup I_b$ into $I_{a\backslash b}$, $I_{b\backslash a}$ and I_{ab} where $I_{a\backslash b}$ is background information that A has but not B, $I_{b\backslash a}$ is background information that B has but not A and I_{ab} is the background information common to A and B. Thus the elements of the set $\{I_{a\backslash b}, I_{b\backslash a}, I_{ab}\}$ are mutually exclusive and their union is I.

Assume now that A is a fact-finder whose background information is formally independent of the scientific evidence E and has no effect on the probability of E and that B is a forensic scientist whose background information is formally independent of the propositions H_p and H_d and has no effect on the probability of H_p or H_d. Then

$$\frac{\Pr(H_p \mid E, I_{a\backslash b}, I_{b\backslash a}, I_{ab})}{\Pr(H_d \mid E, I_{a\backslash b}, I_{b\backslash a}, I_{ab})} = \frac{\Pr(E \mid H_p, I_{a\backslash b}, I_{b\backslash a}, I_{ab})}{\Pr(E \mid H_d, I_{a\backslash b}, I_{b\backslash a}, I_{ab})} \times \frac{\Pr(H_p \mid I_{a\backslash b}, I_{b\backslash a}, I_{ab})}{\Pr(H_d \mid I_{a\backslash b}, I_{b\backslash a}, I_{ab})}$$

$$\Rightarrow \frac{\Pr(H_p \mid E, I_a, I_b)}{\Pr(H_d \mid E, I_a, I_b)} = \frac{\Pr(E \mid H_p, I_b)}{\Pr(E \mid H_d, I_b)} \times \frac{\Pr(H_p \mid I_a)}{\Pr(H_d \mid I_a)}$$

$$\Rightarrow \frac{\Pr(H_p \mid E, I)}{\Pr(H_d \mid E, I)} = \frac{\Pr(E \mid H_p, I_b)}{\Pr(E \mid H_d, I_b)} \times \frac{\Pr(H_p \mid I_a)}{\Pr(H_d \mid I_a)}. \quad (2.17)$$

The reformulation in (2.17) is justified by the assumptions earlier as to the identities of A and B. Thus $\Pr(E \mid H_p, I_{a\backslash b}, I_{b\backslash a}, I_{ab})$ may be written as $\Pr(E \mid H_p, I_b)$ as E is independent of $I_{a\backslash b}$ and dependent only on $I_{b\backslash a} \cup I_{ab} = I_b$. Similarly, $\Pr(E \mid H_d, I_{a\backslash b}, I_{b\backslash a}, I_{ab}) = \Pr(E \mid H_d, I_b)$, $\Pr(H_p \mid I_{a\backslash b}, I_{b\backslash a}, I_{ab}) = \Pr(H_p \mid I_a)$ and $\Pr(H_d \mid I_{a\backslash b}, I_{b\backslash a}, I_{ab}) = \Pr(H_d \mid I_a)$.

The independence argument used for I_a and I_b is formal as participants A and B have information I_{ab} in common. The fair administration of the criminal justice system relies on this common information being treated appropriately.

2.4.5 Summary of Competing Propositions

Initially, when determining the value of evidence, the two competing propositions were taken to be that the PoI is guilty and that the PoI is innocent. However, these are not the only possible propositions as discussed in Chapters 5 and 6.

Care has to be taken in a statistical discussion about the evaluation of evidence as to the purpose of the analysis. Misunderstandings can arise. For example, it has been said that 'the statistician's objective is a final, composite probability of guilt'

(Kind, 1994). This may be an intermediate objective of a statistician on a jury. A statistician in such a position would then have to make a decision as to whether to find the defendant guilty or not guilty. Determination of the guilt, or otherwise, of a defendant is the duty of a judge and / or jury. The objective of a statistician advising a scientist on the value of the scientist's evidence is rather different. That objective is an assessment of the value of the evidence under (at least) two competing propositions. The evidence that is being assessed will often be transfer evidence. The propositions may be guilt or innocence. However, in many cases they will be something else, e.g. the DNA at the crime scene is from the PoI. An illustration of the effect of evidence on the odds in favour of a proposition H_p relative to a proposition H_d has been given in Table 2.6.

However, it has to be emphasised that the determination of the prior odds is also a vital part of the equation. That determination is part of the duty of the judge and / or jury.

Several suggestions for competing propositions, including those of guilt and innocence, are given in the following text and extended in Section 5.2.

(1) H_p: the PoI is guilty;
H_d: the PoI is innocent.

(2) H_p: the PoI stabbed the victim at the crime scene;
H_d: another man stabbed the victim at the crime scene.

(3) H_p: the crime sample came from a Caucasian;
H_d: the crime sample came from an Afro-Caribbean.

(4) H_p: the alleged father is the true father of the child;
H_d: the alleged father is not the true father of the child.

(5) H_p: the two crime samples came from the PoI and one other man;
H_d: the two crime samples came from two other men.

(6) H_p: the PoI was the person who left the crime stain;
H_d: the PoI was not the person who left the crime stain.

(7) H_p: paint on the injured party's vehicle originated from the vehicle of the PoI;
H_d: paint on the injured party's vehicle originated from a random vehicle.

In general, the two propositions can be referred to as the prosecutor's and defence propositions, respectively. The prosecutor's proposition is the one to be used for the determination of the probability in the numerator, the defence proposition is the one to be used for the determination of the probability in the denominator.

At a particular moment in a trial, the context is restricted to (at least) two competing propositions. Consider a rape case in which the victim reported to police that she had been raped by a former

boyfriend. A T-shirt belonging to the boyfriend is examined and foreign fibres are collected from it. The propositions offered may include

H_p: the PoI is the offender;

H_{d1}: the PoI is not the offender and has not seen the victim during the past three weeks;

H_{d2}: the PoI is not the offender but on the night of the alleged rape he was seen dancing with the victim.

The evidence includes the attributes of the foreign fibres found on the boyfriend's T-shirt and the attributes of fibres taken from the victim's garments. The value of this evidence will change as the propositions offered by the prosecution and defence change. Background information, I, has also to be taken into account when defining propositions. Examples are presented in Chapters 5 and 6.

Example 2.7. The following discussion is simplistic in the context of DNA profiling and is provided to illustrate simply the use of the likelihood ratio for the evaluation of evidence. More realistic examples for DNA profiling are given in Section 6.1.5.

A crime has been committed. A bloodstain is found at the scene of the crime. All innocent sources of the stain are eliminated and the criminal is determined as the source of the relevant stain. The purpose of the investigation is to determine the identity of the criminal; it is at present unknown. For the LDLR locus, the bloodstain is

of genotype Γ with a relative frequency γ in the relevant population. A PoI has been identified with the same genotype for locus LDLR as the crime stain. These two facts, the genotype of the crime stain (E_r) and the genotype of the PoI (E_c) together form the observations for the evidence $E = (E_r, E_c)$. Notice again that locus LDLR is no longer used in forensic genetics; it is used here for ease of explanation.

The likelihood ratio can be evaluated as follows. If the PoI is guilty (H_p), the genetical correspondence between the two genotypes is certain and $\Pr(E \mid H_p) = 1$. If the PoI is innocent (H_d), the genetical correspondence is coincidental. The criminal is known to be the source of the crime stain and hence is of genotype Γ. The probability that the PoI would also have group Γ is γ, the occurrence of Γ in the relevant population. Thus, $\Pr(E \mid H_d) = \gamma$.

The likelihood ratio is $\Pr(E \mid H_p) / \Pr(E \mid H_d) = 1/\gamma$. The odds in favour of G are multiplied by $1/\gamma$.

Consider the following numerical illustration of this for the LDLR locus. The three possible genotypes and their frequencies (γ equals the percentage figures of Table 1.1 divided by 100) for a Caucasian Chicago population (Johnson and Peterson, 1999) were given in Table 1.1. The effect ($1/\gamma$) on the odds in favour of H_p is given in Table 2.7 for each genotype.

Verbal interpretations may be given to these figures. For example, if the crime stain were of genotype *AA*, it could be said that 'the evidence

Table 2.7 Value of the evidence for each genotype, given genotypic relative frequencies for locus LDLR amongst Caucasians in Chicago based on a sample of size 200.

Genotype	(Γ)	AA	BB	AB
Value	$(1/\gamma)$	5.32	3.12	2.04

Source: From Johnson and Peterson (1999), Table 1.1. Reprinted with permissions of ASTM International.

of the genotype of the crime stain matching the genotype of the PoI is about five times as likely if the PoI is guilty than if they are innocent'. If the crime stain were of type AB, it could be said that 'the evidence ... is twice as likely if the PoI is guilty than if they are innocent'.

Further discussion of these issues is given in Chapter 5.

2.4.6 Qualitative Scale for the Value of the Evidence

The quantitative value of evidence has been given a qualitative interpretation; examples are Jeffreys (1983), Evett (1987a, 1990), Evett et al. (2000d), Nordgaard et al. (2012a), Marquis et al. (2016). Three examples are presented here, one for posterior odds and two for the likelihood ratio.

Consider two competing propositions H_p and H_d and a value V for a piece of evidence. The European Network of Forensic Science Institutes

guideline (ENFSI, 2015) provides – just for illustrative purposes – an example of a qualitative scale (see Table 2.8).

Note that this type of scale applies also to reciprocal values for $V < 1$ in support of the defence proposition. Also, weak support for the first proposition does not mean strong support for the alternative.

From a historical point of view, a verbal scale for a numerical ratio of probabilities in the context of hypothesis testing was discussed by Jeffreys (1983) (first edition in 1939). A ratio, denoted as K by Jeffreys, is that of the probability of the null hypothesis given evidence and background information to the probability of the alternative hypothesis given evidence and background information. Jeffreys takes the prior probabilities of the two hypotheses to be equal so that the posterior odds equals the likelihood ratio. The verbal summary is then phrased in the form of support provided by the evidence against the null hypothesis. Jeffreys (1983) commented that K need not be known with much accuracy. If $K > 1$ the null hypothesis is supported. Jeffreys further commented that interest is with the values of $K < 1$ when the null hypothesis may be rejected. A logarithmic scale with so-called grades and the associated verbal descriptors proposed by Jeffreys is given in Table 2.9.

Interestingly, Fienberg (1989) provided a quote from a nineteenth-century jurist, Jeremy Bentham, which appeared to anticipate the

Table 2.8 Qualitative scale for reporting the value V of the support of the evidence for H_p against H_d (Source: Based on ENFSI, 2015).

1	$< V \leq$	2	No support
2	$< V \leq$	10	Weak support for the first proposition relative to the alternative
10	$< V \leq$	100	Moderate support for the first proposition relative to the alternative
100	$< V \leq$	1 000	Moderately strong support for the first proposition relative to the alternative
1 000	$< V \leq$	10 000	Strong support for the first proposition relative to the alternative
10 000	$< V \leq$	1 000 000	Very strong support for the first proposition relative to the alternative
1 000 000	$< V$		Extremely strong support for the first proposition relative to the alternative

Table 2.9 Verbal scale of support K for a null hypothesis NH proposed by Jeffreys (1983) where K is the ratio of the probability of the null hypothesis given evidence and background information to the probability of the alternative hypothesis given evidence and background information.

Grade	Value of K	Verbal descriptor
0	$K > 1$	Null hypothesis NH supported
1	$1 > K > 10^{-1/2}$	Evidence against NH but not worth more than a bare mention
2	$10^{-1/2} > K > 10^{-1}$	Evidence against NH substantial
3	$10^{-1} > K > 10^{-3/2}$	Evidence against NH strong
4	$10^{-3/2} > K > 10^{-2}$	Evidence against NH very strong
5	$10^{-2} > K$	Evidence against NH decisive

Source: From Jeffreys, H. (1983). Reprinted with permissions of Oxford University Press.

Jeffreys scale, though perhaps in application to the strength of the belief in the proposition of guilt rather than in the strength of the evidence.

> *The scale being understood to be composed of ten degrees - in the language applied by the French philosophers to thermometers, a decigrade scale - a man says, My persuasion is at 10 or 9* etc. *affirmative, or at least 10* etc. *negative . . . Bentham (1827) quoted in Fienberg (1989, p. 212)*

More recently, an alternative approach is to use an ordinal scale for the likelihood ratio. In forensic

science such a procedure was introduced by Evett (1991) for ease of communication. A similar scale was then proposed by others; examples are presented in Nordgaard et al. (2012a) and in Nordgaard and Rasmusson (2012a) with a neutral response for the likelihood ratio V close to 1 and then a four-point scale for $V > 1$ with its reciprocal for $V < 1$. The example is given in Table 2.10.

In 2015, as illustrated in Table 2.8, a six-point verbal scale was proposed for values of the likelihood ratio greater than 1, with corresponding

Table 2.10 Verbal scale of support for a likelihood ratio V proposed by Nordgaard et al. (2012b) and Nordgaard and Rasmusson (2012a).

Scale	Interval of likelihood ratio V	Degree of support
+4	$10^6 \leq V$	Extremely strong support
+3	$6\,000 \leq V < 10^6$	Strong support
+2	$100 \leq V < 6\,000$	Support
+1	$6 \leq V < 100$	Support to some extent
0	$1/6 \leq V < 6$	Support neither . . . nor
−1	$1/100 \leq V < 1/6$	Support to some extent
−2	$1/6\,000 < V \leq 1/100$	Support
−3	$1/10^6 < V \leq 1/6\,000$	Strong support
−4	$V \leq 1/10^6$	Extremely strong support

reciprocals for values of the likelihood ratio less than 1, and with six adjectives for support of weak, moderate, moderately strong, strong, very strong, and extremely strong and corresponding numerical ranges for the logarithm to base 10 of the likelihood ratio of $\{0.3-1, 1-2, 2-3, 3-4, 4-6, >6\}$.

Another example of the use of a verbal scale can be found in *People v. Carter* (2016) where it is said that the Office of the Chief Medical Examiner (OCME) interprets likelihood ratios as follows: a likelihood ratio of 1.00 is inconclusive; a likelihood ratio in the range of 1.1 to 10 provides limited support for the proposition that the defendant was a contributor; a likelihood ratio in the range of 10 to 100 provides moderate support; a likelihood ratio in the range of 100 to 1000 provides strong support; and a likelihood ratio greater than 1000 provides very strong support of one scenario over the other (p. 2).

For the scientist interested in a qualitative interpretation for the likelihood ratio, one of the scales in Tables 2.8 and 2.10 may be used. Alternatively, the scientist may use a scale of their own devising. Whichever scale is used, it has to be made clear in advance to the fact-finder the scale that is being used.

For DNA in which there can be very large likelihood ratios when considering source level propositions, the verbal scale could appear as inadequate. However, it has become accepted practice since Evett et al. (2000d) to use the

phrase *extremely strong support* for likelihood ratios of one million or more. Note that a comment on extremely high values obtained with DNA evidence is presented in Evett et al. (2000b), Hopwood et al. (2012) and in Section 6.1.5.

Scales based on logarithms (Kass and Raftery, 1995) provide a conversion from the logarithm of prior odds to a logarithm of posterior odds. The use of logarithms transforms the relationship between prior and posterior odds and the likelihood ratio into an additive one. The first suggestion of the use of logarithms to assess the weight of evidence appears to be Peirce (1878) and a useful discussion of this is given in Schum (1994). Thus, from (2.10) with a change of notation

$$\begin{aligned}&\log\{\Pr(H_p|E)/\Pr(H_d|E)\}\\&\quad=\log\{\Pr(E|H_p)/\Pr(E|H_d)\}\\&\qquad+\log\{\Pr(H_p)/\Pr(H_d)\}.\end{aligned}$$

Logarithms provide a good way of comprehending the magnitude of large numbers.

An alternative presentation is

$$\begin{aligned}&\log\{\Pr(H_p \mid E)\}-\log\{\Pr(H_d \mid E)\}\\&\quad=\{\log\Pr(E \mid H_p)+\log\Pr(H_p\}\\&\qquad-\{\log\Pr(E \mid H_d)+\log\Pr(H_d)\}\end{aligned}$$

which illustrates the analogy with the scales of justice. Terms involving H_p go in one tray, terms involving H_d go in the other tray, with the arguments *pro* and *con* of Peirce (1878).

Consider logarithms to base 10. A relative frequency of 1 in 10 million has a logarithm of −7, a relative frequency of 1 in 1 million has a logarithm of −6. The corresponding reciprocals have logarithms of 7 and 6. These numbers, 7 and 6, are much more meaningful to the statistical layman and the difference between them is much more comprehensible. This is because logarithms are also used to measure effects in other areas with which many people are familiar. The Richter scale for measuring the strength of earthquakes is a logarithmic scale. Sound is measured in decibels, another logarithmic scale. The pH scale for measuring acidity is also a logarithmic scale.

As an illustration, consider a case in which the likelihood ratio is 500 million in favour of the prosecution's proposition. Prior odds in favour of innocence of 1 million to 1 will be converted to posterior odds of 500 to 1 in favour of guilt.

In the verbal scale proposed by Aitken and Taroni (1998), analogous to a medical context (Calman, 1996; Calman and Royston, 1997), odds in favour of innocence of 1 million to 1 would relate to a city of 1 million and one people in which there was one guilty person. A person is selected at random from this city. The probability they are the guilty person is approximately 1 in 1 million.

Consider now posterior odds of 500 to 1 in favour of guilt. Imagine a large street with 501 inhabitants. All except one are guilty. A person is selected at random from the street. The probability

he is guilty is 500/501 and the odds are 500 to 1 in favour of guilt.

Different people will have different prior odds and thus different posterior odds. Tables can be provided, such as Table 2.6, to show how the prior odds are converted to posterior odds. Also, from Aitken and Taroni (1998), Table 2.11 converts log prior odds to log posterior odds.

As an example as to how the table may be used consider a case in which the prior odds against the defendant are 1 million to 1, 10 million to 1 or 100 million to 1 with logarithms of -6, -7, -8, respectively. A likelihood ratio of 100 million (with log LR of 8) is obtained. The posterior odds in favour of the guilt of the defendant are then 100 to 1, 10 to 1 or evens (log posterior odds of 2, 1 or 0, respectively).

A final note of the use of verbal qualifiers refers to a potential misunderstanding of their meaning. Notice what is not said in the verbal interpretation. Consider, for the sake of illustration, a blood of profile *AB* for locus LDLR again. It is not claimed that the evidence is such that the PoI is twice as likely to be guilty as they would have been if the evidence had not been presented. It is the evidence that is twice as likely, not the proposition of guilt. The largest value for γ in Table 2.7 is 5.32, a value that would be said to provide weak support using the scale in Table 2.8.

There is no general agreement amongst jurists concerning the association of verbal and numeric scales. Until there is such an agreement, the verbal

Table 2.11 The values of the logarithm of the posterior odds in favour of an issue determined from the values of the logarithm of the prior odds in favour of guilt (log (Prior odds)) and the logarithm of the likelihood ratio.

Verbal description	Log (prior odds)	Logarithm of the likelihood ratio												
		−2	−1	0	1	2	3	4	5	6	7	8	9	10
Individual	0	−2	−1	0	1	2	3	4	5	6	7	8	9	10
Family	−1	−3	−2	−1	0	1	2	3	4	5	6	7	8	9
Street	−2	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8
Village	−3	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
Small town	−4	−6	−5	−4	−3	2	−1	0	1	2	3	4	5	6
Large town	−5	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5
City	−6	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4
Province / country	−7	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3
Large country	−8	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2
Continent	−9	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1
World	−10	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0

The values in the body of the table are obtained by adding the appropriate row and column values. Logarithms are taken to base 10. The verbal description is taken from Calman and Royston (1997). Source: Based on Calman and Royston (1997).

description for a numeric value will have to remain a matter of personal judgement.

This consideration of support for a proposition is illustrated in the following case from 1998.

> *We note and we follow and accept unreservedly Dr Evett's evidence to us and his strictures to us that we cannot look at one hypothesis, we must look at two and we must test each against the other … what is the probability of the evidence if the Respondent's hypothesis is correct? what is the probability of the evidence if the Appellant's hypothesis is correct?*
>
> *Dr Evett tells us (and we follow it) that if the answer to the first question is greater than the answer to the second question, then the Respondent's hypothesis is supported by the evidence.*
>
> *(*Johannes Pruijsen vs. H.M. Customs & Excise.*)*

Statements concerning the probability of guilt require knowledge of the prior odds in favour of guilt, something that is not part of the scientist's knowledge.

A similar interpretation is given in Royall (1997)

> *[. . .] a likelihood ratio of k corresponds to evidence strong enough to cause a k-fold increase in a prior probability ratio, regardless of whether a prior ratio is actually available in a specific problem or not. (p. 13)*

Note that this is a measure of the value of the evidence. The implications for a particular value of evidence will vary according to the context.

Finally, note that whilst it is permissible to interpret a numerical value verbally, it is not meaningful to interpret a verbal scale numerically.

Also, there are still several aspects of the use of verbal scales to be considered. First, there is the nature of the assistance that a verbal scale might offer to the fact-finder (judge or jury). Comments are presented in Martire et al. (2014) and Martire and Watkins (2015) who collected evidence supporting the proposition that the verbal formulations of uncertainty are a less effective form of communication than the equivalent numerical formulations. The second is whether the numerical value of a likelihood ratio is a sufficient summary of the value of the evidence. In that respect, Meester and Sjerps (2004) noted that one should be careful when using expression like 'the evidence is more likely under H_p than under H_d' if the propositions were suggested by the data. They (Meester et al., 2006) supported that 'in that case, they only become meaningful in combination with prior probabilities for the proposition considered' (p. 245). Thirdly, it should be noticed that the choice of a verbal scale is initially subjective but not arbitrary. However, if the scale is to have credibility and be acceptable to the courts, then a particular choice has to be agreed amongst the scientists across all forensic disciplines covered by a given laboratory. So, as mentioned in the ENFSI (2015) guideline, given that the purpose of such a scale is to assist the court in relation to the value of the findings, 'it is incorrect to use different scales for different types of evidence (e.g. DNA and glass).' (p. 18). Fourthly, it has been pointed out that the perceived strength of several types of

statements on evidence evaluation relative to one another is modified relative to persons. Thompson et al. (2018) questioned if particular statements convey the intended meaning and are interpreted in a justifiable and appropriate way. Finally, the limitations of the use of verbal scales need to be recognised. For example, verbal qualifiers cannot be coherently combined with other evidence. A discussion of the disadvantages of verbal scales is to be found in Marquis et al. (2016) and in Berger and Stoel (2018).

Marquis et al. (2016) conclude by affirming:

> *Finally, we are of the opinion that if experts were able to coherently articulate numbers, and if the recipients of information could properly handle such numbers, then verbal qualifiers could be abandoned completely. (p. 364)*

Discussions on perception problems relate to verbal statements can be found in Martire et al. (2013), Mullen et al. (2014), and Arscott et al. (2017).

2.5 ERRORS IN INTERPRETATION

A large part of the controversy over scientific evidence is due to the way in which the evidence has been classically presented. At trial, it is already a complex operation to ensure that judge and jury members understand the scientific evidence; additional difficulties are added when the forensic scientist gives the court an evaluation

to illustrate the convincing force of the results (see, for example, the misunderstanding in *R v. Adams, D.J.*(1997). Kirk and Kingston (1964) described the problem by affirming that

> *The first problem centres on the idea of* certainty. *The witness is often very certain that he is right. Where this degree of certainty exists in fact, there is no need for statistics. However, if an opposing and presumably equivalent witness is equally certain of a different interpretation, we are confronted with a serious impasse. One or the other or both must be wrong to some extent. Here, obviously, the first remedy is for each to abandon the idea of absolute certainty. […] If it can be accepted that nothing is absolutely certain, then it becomes logical to determine the degree of confidence that may be assigned to a particular belief. It is here that statistics offer the most valuable approach. (p. 435)*

The assessment of the value of the analytical results is associated with probabilities as measures of uncertainty. Hence experts' statements are associated with uncertainty. It is important to ensure that this uncertainty is measured accurately and represented correctly to avoid the so-called fallacies or pitfalls of intuition (Saks and Koehler, 1991; Fienberg and Kaye, 1991; Eggleston, 1991; Kaye, 2015). Kirk and Kingston (1964) affirmed that 'The development of a firm statistical foundation could provide a court of appeals for the disagreeing witnesses.' (p. 437).

Psychological researches have demonstrated that intuition is a bad substitute for the laws of probability in evaluating uncertainty (Bar-Hillel and Falk, 1982; Tversky and Kahneman, 1982;

Koehler, 1992; Piattelli-Palmarini, 1994; Kahneman, 2011), and the presentation of scientific argument at trial can create confusion (Zeisel and Kaye, 1997). Victims of this confusion are both jurists and experts (Koehler, 1993b; Reinstein, 1996).

More recent research focuses on the relationship between the ability to make correct probabilistic evaluations and individuals' culture and schooling; see, for example, Fontanari et al. (2014), Girotto and Pighin (2015), Pighin et al. (2017).

Statistical reasoning will support the forensic scientists and members of the jury in reaching conclusions. Since the beginning of the twentieth century, some scientists and jurists were plainly conscious of the lack of intuition in dealing with the calculation of chances. The French mathematician Henri Poincaré gave a remarkable example of these limits in his course taught during 1895 under the title 'The problem of the three caskets' (Poincaré, 1896), also known as the 'Monty Hall problem'. Around 100 years later, the same problem creates large controversies (Selvin, 1975; Engel and Ventoulias, 1991; Morgan et al., 1991; Falk, 1992).

Consider three boxes A, B, and C, within one of which there is a prize. There is a competitor, K, and a compère, L. The compère knows the identity of the box which contains the prize. The prior probability $\Pr(i)$ that a particular box contains the prize is $1/3$ for $i = A, B, C$. K has to choose a box. If the chosen box contains the prize, K keeps the

prize. The choice is made in two stages. K chooses a box but it is not opened. L then opens one of the two other boxes and shows K that it is empty (remember L knows which box contains the prize). K is then allowed to remain with the first choice of box or to change to the other unopened box. What should K do in order to maximise the probability of winning the prize?

Suppose, without loss of generality, that the competitor initially chooses C. The compère, L, opens one of A or B, which is empty. Assume, again without loss of generality, that A is opened. If the prize is in C, then $\Pr(L \text{ opens } A) = \Pr(L \text{ opens } B) = 1/2$. If the prize is in B, then $\Pr(L \text{ opens } A) = 1$. If the prize is in A, then $\Pr(L \text{ opens } A) = 0$.

The probability of interest is $\Pr(\text{true box is } C \mid L \text{ opens } A \text{ and } K \text{ chooses } C)$. Denote this as $\Pr(C \mid L_A, K_C)$. As K_C is common to what follows, it will be omitted for ease of notation. Let L_B and L_C denote the events that the compère opens B and C, respectively. The probability of interest $\Pr(C \mid L_A)$ may then be written as

$$\frac{\Pr(L_A \mid C)\Pr(C)}{\begin{array}{c}\Pr(L_A \mid A)\Pr(A) + \Pr(L_A \mid B)\Pr(B) \\ + \Pr(L_A \mid C)\Pr(C)\end{array}} \tag{2x}$$

which is equal to

$$\frac{\frac{1}{2} \times \frac{1}{3}}{0 + (1 \times \frac{1}{3}) + (\frac{1}{2} \times \frac{1}{3})} = \frac{1}{3}.$$

The prior probabilities for the box that contained the prize were $\Pr(A) = \Pr(B) = \Pr(C) = 1/3$. The competitor chose C. The compère then opened A and showed that it did not contain the prize. The calculations earlier (2x) show that the posterior probability that C contains the prize is still $1/3$. However, A has been eliminated. Thus, the posterior probability that B contains the prize is $2/3$. The odds are 2 to 1 in favour of B containing the prize. The competitor should thus change his choice of box.

The solution is therefore that changing will let one win twice as often as keeping the original choice; such a solution seems counter-intuitive. Note logically that in two out of three scenarios, you win by changing your selection after one of the doors is revealed. This is because there is a greater probability one chooses a box without the prize in their first choice and so the compère is guaranteed to reveal one of the other boxes that has no prize. By subsequently changing your first choice, you double your probability of winning the prize. Another interesting discussion about this and other probabilistic conundrums can be found in D'Agostini (2010).

More subtle forms of pitfalls of intuition have been described in the analysis of court reports and scientists' statements as evidence. As specified by Kingston (1966)

> *[i]t must be realised that probability, and consequently probability theory* does *have a place in court. It is difficult to imagine that any reasonable assessment of circumstantial evidence can be made without the elementary principles*

> *of probability theory being used, whether this use is made explicit or not. Generally the basic probabilistic ideas remain unexpressed, and serve only as mental pathways guiding the jurors' thought processes to their conclusions. But every so often these ideas are made explicit and are cast into mathematical terms. When this occurs, it is essential that the manipulation and interpretation of these mathematical terms be correct. Unfortunately, in many reported instances where probability figures have been expressed in criminal trials, or discussed in connection with such proceedings, gross errors have been made. (p. 93)*

As an example of the pitfalls of intuition, case reports on DNA evidence were studied and submitted to practitioners (forensic scientists, lawyers, advocates) and students to investigate their understanding of measures of uncertainty (Taroni and Aitken, 1998b,c). Results suggested improvements for the presentation of scientific evidence and for the education of future lawyers and forensic scientists. Examples of these fallacies abound in literature since 1990 and are presented in the following sections 2.5.1–2.5.8 (see also Goodman 1992; the historical criminal case *R v. Deen*, *The Times*, January 10, 1994; Matthews 1994; Dickson 1994; Balding and Donnelly 1994).

More recently, an extended series of guidance texts for judges, lawyers, forensic scientists, and expert witnesses have been published by the Royal Statistical Society under the general theme of 'Communicating and interpreting statistical evidence in the administration of criminal justice'. The first guidance (entitled 'Fundamentals of

probability and statistical evidence in criminal proceedings') mentioned logical problems linked to probabilistic reasoning and the confusion between expressions for the likelihood ratio and posterior odds and more generally amongst conditional probabilities (Aitken et al., 2010). Such confusions were also detected in other scientific fields (see, for example, D'Agostini (2003), Crupi et al. (2018)). Note that forensic literature also emphasised problems generated by the misunderstanding of probabilities and noted that (Berger et al., 2011) 'the judgement will be interpreted as being in opposition to the principles of logical interpretation of evidence' (p. 43). An official scientific report (US President's Council of Advisors on Science and Technology (PCAST, 2016)), has been criticised for errors related to pitfalls of intuition in the probabilistic domain. Evett et al. (2017) wrote:

> *The most serious weakness in the PCAST report is their flawed paradigm for forensic evaluation. Unfortunately, the report contains more misconceptions, fallacies, confusions and improper wording. (p. 18)*

2.5.1 Fallacy of the Transposed Conditional

Examples of this fallacy abound in judicial and forensic literature. References to it in the courts began with Bertillon's testimony in the Dreyfus case and include cases since 1990 *R. v. Adams,D.J.*

(1997), R. v. Doheny and Adams,G. (1997), R. v. Clark (2003), R. v. T. (2010), and Wilson v. Maryland (2002) as presented in Dawid (2004) and in Section 3.2.

Consider the following example, from Gaudette and Keeping (1974). The authors conducted a lengthy experiment to attempt to determine the ability of scientists to distinguish between different people on the basis of an interpretation of hair samples. Multiple comparisons were made of hairs from many different people. In one experiment, nine hairs, selected to be mutually dissimilar, from one source were compared, one by one, with a single hair from another source. It was estimated, from the results of many such comparisons, that the probability that, in at least one of the nine comparisons, the two hairs examined, from different sources would be indistinguishable would be 1/4500. The authors concluded that 'it is estimated that if one human scalp hair found at the scene of a crime is indistinguishable from at least one of a group of about nine dissimilar hairs from a given source the probability that it could have originated from another source is very small, about 1 in 4500' (Gaudette and Keeping, 1974, p. 605).

Let R denote the event that 'one human scalp hair found at the scene of a crime is indistinguishable from at least one of a group of about nine dissimilar hairs from a given source'. Let S denote the event that 'the nine dissimilar hairs come from a different source than the single hair'. Then, the

authors' experiment gives a value for $\Pr(R \mid S)$ but the authors' summarising statement gives a value for $\Pr(S \mid R)$.

Another example is the Dreyfus case, more details of which are in Section 3.2 and the probabilistic testimony offered by Alphonse Bertillon. Bertillon failed with the same problem of intuition (for a full description of the judicial case and comments on experts' conclusions, see Champod et al. (1999)). According to Bertillon, Dreyfus was the author of a so-called bordereau (document). To increase the credibility of his allegations, Bertillon submitted a probabilistic calculation. If the individual probability for one coincidence were set to 0.2, then the probability of observing four coincidences would be $0.2^4 = 0.0016$ and generally for N coincidences, the probability would be 0.2^N. Considering the four coincidences observed by Bertillon, the probability 0.0016 was considered so remote that it demonstrated the forgery. Even if it is admitted that the value $p = 0.0016$ provided by Bertillon was correct (for a comment on this point, see Darboux et al. (1908), Champod et al. (1999)), he claimed (indirectly) that it was possible to deduce from p that the probability that the questioned document was a forgery was $1 - p$. This latter probability was sufficiently close to 1 that it constituted an unequivocal demonstration with a reasonable degree of scientific certainty that Dreyfus was the author. Bertillon's statement is fallacious because he seemed to argue that $p = \Pr(H_d \mid E)$, hence

$\Pr(H_p \mid E) = 1 - p$, whereas p only represents $\Pr(E \mid H_d)$.

Other examples of the fallacy of the transposed conditional are given by Thompson and Schumann (1987) who gave it the name of the *prosecutor's fallacy*. It has also been called the *inversion fallacy* (Kaye and Koehler, 1991; Kaye, 1993). For example:

> *There is a 10% chance that the defendant would have the crime blood type if he were innocent. Thus there is a 90% chance that he is guilty.*

or

> *The blood test is highly relevant. The suspect has the same blood type as the attacker. This blood type is found in only 1% of the population so there is only a 1% chance that the blood found at the scene came from someone other than the PoI. Since there is a 1% chance that someone else committed the crime there is a 99% chance that the suspect is guilty. (Thompson and Schumann, 1987, p. 177).*

In general, let E denote the evidence and H_d the proposition that a suspect is innocent. A value is determined for $\Pr(E \mid H_d)$, the probability of the evidence if the PoI is innocent.

The ENFSI Guideline for evaluative reporting (ENFSI, 2015) simply describes the occurrence of such a pitfall by saying:

> *In the legal context, a fallacious transposed conditional statement is one that equates (or, confuses) the probability of particular findings given a proposition with the probability of that proposition given these findings. (p. 27)*

Such a sentence reiterates what was listed as a problem in a 2011 guest editorial (Association of Forensic Science Providers, 2011) and in Berger et al. (2011):

> *It is necessary for the scientist to consider the* probability of the observations *given each of the stated propositions. Not only it is not appropriate for the scientist to consider the* probability of the proposition *given the observations, there is a danger that in doing so the jury will be misled. (p. 1)*

The interpretation of the value calculated can cause considerable confusion. Two special cases of the fallacy of the transposed conditional in which $\Pr(E \mid H_d)$ is confused with

(a) the probability the PoI is not the source of the evidence; known as the *source probability error*;

(b) the probability the PoI is not guilty; known as the *ultimate issue error*;

are discussed by Koehler (1993a), Evett (1995), Redmayne (1995, 1997).

2.5.2 Source Probability Error

A crime is committed. Trace evidence is found, which is considered to have come from the criminal. Let H_d be the proposition that the PoI was not the source of the evidence.

For instance, the evidence E may be that a DNA match has been found between blood from a murder victim and blood recovered from the clothing

of a PoI. A scientist determines a value for Pr(E | H_d) as 1 in 7 million. Consider the following possible statement of the value of the evidence (based on *Wike v. State* (1992), transcript, pp. 147–148, given in Koehler (1993a)).

> *With probability 1 in 7 million, it can be said that the blood on the clothing of the suspect could be that of someone other than the victim.*

Other possibilities are, where the figures quoted are for illustrative purposes only:

- ... the probability that the DNA that was found at the scene of the crime came from anyone else other than the PoI is 1 in 7 million ...;
- ... the probability of this DNA profile occurring at random is 1 in 18 billion; thus the likelihood that the DNA belongs to someone other than the PoI is 1 in 18 billion ...;
- ... the probability of finding the evidence on an innocent person is 0.01% (1 in 10 000) thus the likelihood that the suspect is guilty is 99.99% ...;
- ... the trace evidence has the same DNA profile as the PoI, thus the trace evidence has been left by the PoI ...;
- After conducting DNA testing on the vaginal swab samples taken from the victim and Ross' [the PoI] blood samples, the DNA expert stated that Ross was the source of the seminal fluid (*E. Ross v. State of Indiana*, 1996).

- The expert offered probability statistics on whether the DNA sample found on the victim came from someone other than defendant (*State of Vermont v. T. Streich*, 1995).

None of the conclusions in these statements is justified by the value given by $\Pr(E \mid H_d)$. All give a mistaken probability for the source of the evidence.

Consider also situations involving the value for the likelihood ratio that is mistakenly interpreted. In *People v. Carter* (2016), at p. 2, it can be read:

> *At issue in this case is the OCME's determination that the DNA mixture found on the sweatshirt was 5640 times more likely to have come from defendant and two other unknown individuals rather than from three unknown individuals.*

It is of interest how the forensic scientist falls into this logical trap after a sequence of questions asked by prosecutors. Examples are the following:

Q And you are able to compile all four of those probabilities and determine what is the likelihood of the DNA found in Billy Glover just randomly occurring in some other DNA sample?

A Yes.

Q What is the likelihood of that?

A The way that is done is to multiply each one of those four numbers that I mentioned before together, because each one is separate and independent, and the final number comes out as one in about 18 billion.

Q So the likelihood that DNA belongs to someone other than Billy Glover is one in 18 billion?

A That is correct.
(*State v. Glover*, 1992),

and

Q Whose blood was found to be on item 52?

A Mr Davis's blood.
(*State v. Davis*, 1991).

The answer of the last expert does not represent an inference on the source, but a decision since there is no probabilistic component to the statement. More on how to make a decision is presented in Section 2.8.

Source probability errors and numerical conversion errors (Section 2.5.6) also occur when experts or jurists try to summarise the likelihood ratio expression. Here is an example (Vincent, 2010):

> *It was 800 billion times more likely that the sample originated from the accused rather than an unknown person. In other words, it would appear to be necessary to search well beyond this planet and conceivably this galaxy to find a match. (p. 40)*

The first sentence is an example of the source probability error. The second sentence is an example of the numerical conversion error (see Section 2.5.6).

Kingston and Kirk (1964) were already aware of such a pitfall of intuition. They wrote:

> *As an example, consider the opinion that a particular suspect left the partial fingerprint in evidence, the basis being a coincidence of 12 elements with no observable differences in the patterns of the evidential print and an area of the print taken from the suspect. What logic leads from the basis of the opinion? Even though the opinion is considered by the courts as an acceptable one from the given basis, there is some disagreement on the connecting logic. (p. 520)*

2.5.3 Ultimate Issue Error

The source probability error may be extended to an error known as the *ultimate issue error* (Koehler, 1993a). This extends the hypothesis that the PoI is the source of the evidence to the hypothesis that the PoI is guilty. The case of *People v. Collins* (1968) (Section 3.4) is a particular example of this. Consider a case in which $\Pr(E \mid H_d)$ is 1 in 5 million, say, where, as before, H_d is the proposition that the PoI was not the source of the evidence E. The ultimate issue error would interpret this as a probability of 1 in 5 million that the PoI was innocent.

2.5.4 Defence Attorney's Fallacy

As well as the fallacy of the transposed conditional, there is also a *defence attorney's fallacy* (Thompson and Schumann, 1987). Consider the following illustrative statement from a defence lawyer

> *The evidence for blood types has very little relevance for this case. Only 1% of the population has the rare blood*

> *type found at the crime scene and in the suspect. However, in a city, like this one in which the crime occurred, with a population of 200 000 people who may have committed the crime this blood type would be found in approximately 2 000 people. The evidence merely shows that the PoI is one of 2 000 people in the city who might have committed the crime. The blood test evidence has provided a probability of guilt of 1 in 2 000. Such a small probability has little relevance for proving the suspect is guilty. (p. 177)*

Strictly speaking (from an inferential point of view) the defence lawyer is correct. However, before the evidence of the blood test was available, the PoI had a probability of only 1 in 200 000 of being guilty (not accounting for any other evidence which may have been presented). The effect of the blood test evidence is to increase this probability by a factor of 100. The evidence is 100 times more likely if the PoI is *guilty* than if they are *innocent*, assuming a probability of 1 for the numerator. This may be thought to show that the blood test evidence is compelling in support of a hypothesis of guilt. Of course this evidence is unlikely to be sufficient on its own. for a verdict of guilty to be returned. This is so because the prior probability of guilt is so small that the resulting posterior probability of guilt will be insufficient for a verdict of guilty to be returned.

As discussed, there are practical difficulties surrounding the use of probabilities to evaluate evidence. It is not uncommon for a court to be faced with two opposing arguments, say, the transposed conditional on one side, and the defender's

line of reasoning on the other side. Mode (1963) wrote:

> *Assume, for example, that the prosecuting attorney has correctly computed the probability to be one in a million that a man selected at random has all of certain characteristics identified with the true criminal, whoever he may be, and concludes that since the accused has these characteristics he must be guilty. The defending lawyer may disparage this evidence by stating that with (say) forty million adult males in the United States and one in a million having the characteristics named, there must be forty men in the country eligible for suspicion. (p. 629)*

Two other errors discussed by Koehler (1993a) are the *probability (another match) error* and the *numerical conversion error*.

2.5.5 Probability (Another Match) Error

As in Example 1.1 (Section 1.3.2) a crime is committed. Evidence E of a blood stain with profile Γ is found at the scene and identified as belonging to the criminal. A PoI is identified. Let H_d be the proposition that the evidence E was not left by the PoI. Suppose the occurrence of the profile of the stain is γ amongst the relevant population (details about such an estimate are given in Section 6.1.5) and so $\Pr(E \mid H_d) = \gamma$. Then the probability that a person selected at random from the population does not match this profile is $(1 - \gamma)$. Let N be the size of the population. Assume, for the purpose of discussion, independence amongst members

of the population with respect to E. Then the probability of no matches with the crime stain profile amongst the N members is $(1-\gamma)^N$ (a generalisation to N events of the third law of probability for independent events, (1.10)). The complement of no matches is at least one match. Hence, the probability of at least one match is $\theta = 1-(1-\gamma)^N$. Two numerical examples are given in Table 2.12. Let $N = 1$ million and take γ as a value that has been estimated from some larger population, assumed similar to the relevant population with regard to profile random match probabilities (see Section 6.1.5). As in Section 4.3.1, see Smith and Charrow (1975) and Finney (1977) for comments about *super-populations*.

Thus it is possible for γ to be less than $1/N$. The two probabilities for θ in Table 2.12 are considerably larger than the corresponding values for γ. The probability (another match) error arises when the two probabilities, γ and θ, are equated. In other words, a small value of γ is taken to imply a small value for the probability that at least one other person has the same matching evidence. The results in Table 2.12 illustrate why this implication is false. A random match probability (RMP)

Table 2.12 Probability θ of at least one match, given a relative frequency of the trace evidence of γ, in a population of size 1 million.

γ:	1/1 million	1/10 million
θ:	0.632	0.095

of 1/1 million (if it is assumed that the value is derived correctly and that there is no possibility of error, lying, or misinterpretation of the data), means that there is 1 chance in 1 million that a single randomly selected person would share the observed characteristics. In other words, assuming that the data and its interpretation are infallible, we would expect to see this DNA profile in approximately 1 out of every million people. Notice that this is not identical to the probability that there exists someone else who shares the observed profile. Although it may be extremely unlikely that a single randomly selected person would share a DNA profile with another person, it may be quite likely that others share this profile (Koehler, 1996).

There is only 1 chance in 1 million that a random person shares a DNA profile that is common to one in every million people, but there is a 63.2% chance that there is at least one other person in a population of size 1 million people (see Table 2.12) who share the profile.

Kingston (1965a) discussed this specific aspect of the probabilistic reasoning. In connection with his analysis, he referred to Galton's 1892 book on fingerprints. Galton reasoned correctly and avoided the fallacy. Kingston wrote:

> *[...] the chance of lineations constructed by the imagination according to strictly natural forms, which should be found to resemble those of a single finger print in all their minutiae, is less than [...] 1 to about sixty-four thousand*

millions. The inference is, that as the number of the human race is reckoned at about sixteen thousand millions, it is a smaller chance than 1 to 4 that the print of a single finger of any given person would be exactly like that of the same finger of any other member of the human race. (p. 75)

2.5.6 Numerical Conversion Error

Let γ be the RMP (see details in Section 6.1.5) of the crime stain as in Section 2.5.5. A match between the crime stain and the profile of a PoI has been made. Let n be the number of people who would have to be tested before there is another match. It may be thought that the significance of the value of γ can be measured by equating $1/\gamma$ with n. A small value of γ implies a large value of n. It is fairly straightforward to calculate n, given γ and given some value for the probability of another match occurring, $\Pr(M)$, say. The numerical conversion error claims that n equals $1/\gamma$ but this is not so.

Suppose $\gamma = 0.01$. There is a probability of 0.01 that a randomly selected individual would match the evidence E. The numerical conversion error would claim that 100 people need to be tested before there is another match but this is not the case. The error is exposed by consideration of $\Pr(M)$. Suppose, initially, that $\Pr(M)$ is taken to be equal to 0.5. A value of n greater than the value determined using a value of $\Pr(M)$ of 0.5 would imply that a match is more likely to happen than not if n people were tested.

Let n be the number of people who are to be tested before the probability of finding a match is greater than 0.5. The probability that a randomly selected individual does not match the evidence is $(1-\gamma)$. For n independent randomly selected individuals, the probability none matches the evidence is $(1-\gamma)^n$. The probability there is at least one match is thus $1-(1-\gamma)^n$. Thus, for a match to be more likely than not with n individuals, $1-(1-\gamma)^n$ has to be greater than 0.5 and so

$$(1-\gamma)^n < 0.5.$$

This inequality may then be written as $n \log(1-\gamma) < \log 0.5$. Thus, $n > \log 0.5/\log(1-\gamma) = \psi_5$, say (remembering that here $(1-\gamma)$ is less than 1 and so its logarithm is negative). A similar argument shows that if $\Pr(M)$ is taken to be greater than 0.9 then $n > \log 0.1/\log(1-\gamma) = \psi_9$, say. Values of ψ and n are given in Table 2.13 for $\gamma = 0.1$, 0.01 and 0.001 and for values of $\Pr(M)$ equal to 0.5 and 0.9.

It is also worth noting that if $n' = 1/\gamma$ people were tested this does not make the probability of a match certain. If n' people are tested, the probability of at least one match is $\theta' = 1-(1-\gamma)^{n'}$; see Table 2.14 for examples. Notice that as $\gamma(=1/n') \to 0$, $\theta' \to 1-e^{-1} = 0.632\ldots$. Notice also that $n_5 < n'$. Thus the numerical conversion error, based on $\Pr(M) = 0.5$, exaggerates the number of people that need to be tested before a match may be expected. For illustrative purposes,

Table 2.13 Evidence occurs with *RMP* γ.

	$\Pr(M) = 0.5$		$\Pr(M) = 0.9$	
γ	ψ_5	n_5	ψ_9	n_9
0.1	6.6	7	21.9	22
0.01	69.0	69	229.1	230
0.001	692.8	693	2 301.4	2 302

Smallest number ψ of people to be observed before a match with the evidence occurs with a given probability, $\Pr(M) = 0.5, 0.9$; $\psi_5 = \log 0.5/\log(1-\gamma)$, $\psi_9 = \log 0.1/\log(1-\gamma)$, n_5 is the smallest integer greater than ψ_5, n_9 is the smallest integer greater than ψ_9.

Table 2.14 The probability, θ', of at least one match with the evidence, which occurs with *RMP* γ, when $n' = 1/\gamma$ people are tested.

γ	n'	θ'
0.1	10	0.65
0.01	100	0.63
0.001	1 000	0.63

consider the following case: The RMP equates 1 in 209 100 000 and the expert said that he has a database of blood samples from all over the country and he asked the question 'How many people would we have to look at before we saw another person like this ?' The answer given is 209 100 000 (*Ross vs. State*, 1992). This exaggerates the probative strength of a match and favours the prosecution.

It is of interest to recall a recent example of such a fallacy. In the 1 July 2016 judgement from the Court of Appeal of Bergamo (Italy) in the Yara Gambirasio case, at p. 96, one can read the DNA expert's statement on DNA evidence evaluation.

> *[. . .] a relative frequency of* 2.33×10^{-27} *is equivalent to certainty on the source of the stain. Given a world population of approximately 7 billion people, you have to imagine 130 millions of billions planets as Earth to find another individual sharing the same genetic characteristics as [the accused person].*[9]

2.5.7 False Positive Fallacy

Consider the possibility of a misclassification error for a scientific sample so that the sample is classified as positive when in fact it is negative. An example of this would be the misclassification as a match, two DNA profiles from different sources that did not match. Serious errors of interpretation can occur through ignorance or underestimation of the potential for a false positive. A low value for the probability of a false positive does not imply that the probability of a false match is low in every

[9]The original Italian text is the following: '[...] con una ricorrenza statistica di 2.33×10^{-27}, equivalente alla certezza. Stimata in sette miliardi la popolazione mondiale, per trovare un altro individuo, oltre a [nome dell'imputato], con le stesse caratteristiche genetiche sarebbero necessari centotrenta milioni di miliardi di altri mondi uguali al nostro [...]'. Note that 130 millions of billions planets similar to Earth in population size generate a figure of 9.1×10^{-26}.

case. A forensic scientist who thinks there is only a probability of 0.01 of declaring, falsely, a match between the samples in a case if they really do not match may assume that there is, necessarily, a probability of 0.99 that the reported match is a true match.

Let M be the event that the PoI and the perpetrator have matching DNA profiles and R be the report of a match. The false positive probability is $\Pr(R \mid \bar{M})$, the probability of reporting a match when the samples do not have matching profiles. The probability $\Pr(M \mid R)$ is the probability of a true match, given that a true match has been reported. The fallacy of the false positive is to equate $\Pr(M \mid R)$ with $1 - \Pr(R \mid \bar{M})$.

The fallacy is a version of the prosecutor's fallacy. Further details are given in Sections 2.7 and 6.1.6.4, and in Thompson et al. (2003).

2.5.8 Expected Value Fallacy

Kaye (2011) – referring to *State v. Wright* (2011) – considered a sex-related crime where it is assumed that the evidence (DNA on the PoI) is a recovered DNA mixture of two persons: the PoI (as the major contributor) and the victim (as a minor contributor). The profile Γ of the minor contributor has a conditional match probability $\gamma = 1/500\,000$. The forensic scientist thought that the conditional match probability was so small that only two people in the state would contribute to this mixture, the PoI and the victim.

Kaye (2011) suggest that the forensic scientist apparently reasoned as follows:

(1) The state population is about 1 000 000;
(2) About 500 000 are women;
(3) The expected number of women in the state with the minor DNA profile is 1 (this value is obtained by multiplying the conditional match probability of 1/500 000 by the female population in the state of 500 000);
(4) No one but the victim could have contributed the minor DNA profile.

The transition from point 3 to point 4 assumes an expected value of 1 for the number of matching profiles in the female population and so it is impossible another matching profile exists in the population. The acceptance of this transition has been called by Kaye (2011), the *expected value fallacy*:

> *The fallacy consists in treating the expected value of a random variable as if it were the only plausible value. (p. 4)*

At point 3, the reported quantity of 1 represents an expected value of a variable about the number of women with the minor DNA profile in a randomly generated population of 500 000 women.

There is a probability associated with the observation of $0, 1, 2, \ldots$ women in a population. In fact, imagine the generation of many populations,

each of size 500 000, in some you can find zero women sharing the given DNA profile, in another you can find one or more. The probability distribution associated with the corresponding random variable X, parametrised by the mean λ, is the Poisson distribution (see Section 7.2.1 for other examples in forensic science and Section A.2.6 for a technical description of the Poisson distribution):

$$\Pr(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!},\ x = 0, 1, \ldots.$$

Given that λ equals 1 (the mean of the occurrence of a given DNA profile, in that case $\gamma \times$ the population size), the probability of observing zero occurrences of such a DNA profile equals $e^{-1}/0! = 0.368$. The probability to observe one occurrence equals $e^{-1}/1! = 0.368$, also.

Given that a DNA profile identical to that of the victim has been observed in the population, a scientist is interested in the probability of observing two or more women with that profile, given that there is at least one woman with the profile. This probability is one minus the probability of observing 0 or 1 women, or $1 - 0.368 - 0.368 = 0.264$ divided by the probability at least one woman has the profile. The forensic scientist erred in saying that the DNA profile must be that of the victim. The probability that such a profile came from more than one woman given at least one woman has the profile is $0.264/(1 - 0.368) = 0.418$.

2.5.9 Uniqueness

Statements of identification (or 'individualization') invoking notions such as 'uniqueness' or 'individuality' are regularly commented on in forensic science literature since Kirk (1963) (e.g. Champod (2000), Saks and Koehler (2008), Champod (2009), Page et al. (2011), Mnookin (2008), Margot (2011)). Such statements are used daily in fields such as fingerprints, shoeprints, toolmarks, firearms, earprints, and speaker recognition (e.g. Simons (1997)). DNA evidence is not immune from this tendency (Kaye (1997), Zeisel and Kaye (1997), Robertson and Vignaux (1998), Balding (1999), and Robertson et al. (2016)).

Kaye (2009a) clarified the terms identification, individualization, and uniqueness and discussed the relationships amongst them. The topic is developed further in Kaye (2013). Cole (2009) argued that trace evidence disciplines do not need these concepts. He explored what defensible conclusions might look like and how they might be supported. He reaffirmed and extended his perspective in Cole (2014).

In the late 1990s, the Federal Bureau of Investigation (FBI) announced that their experts would be permitted to testify that DNA from blood, semen, or other biological evidence recovered at a scene of crime came from a specific person (Holden, 1997). Note, however, that affirming that material recovered at a scene of crime comes from a particular person is an expression on a

proposition and constitutes a so-called transposed conditional statement (see Section 2.5.1). The policy is based on particular statistical figures (e.g. 1 in 260 million) and has been reaffirmed repeatedly since then (Budowle et al. (2000); Moretti and Budowle (2017)). For a critical discussion on the pertinence of the statistics presented in these policies, see, for example, Evett and Weir (1998), Buckleton (2004), and Saks and Koehler (2008) who wrote:

> *Although markers that rarely occur* might *be unique, it is a fallacy to infer uniqueness from profile frequencies simply because they are smaller than the number of available objects. (p. 204)*

An often invoked, but flawed, underlying argument in attempts to justify identification as a concept is that a low, or very low, probability for the event of encountering a given finding in another person (or object) from the relevant population entitles one to jump to the conclusion that such an observation is actually impossible, and hence to claim that a given state of nature (e.g. common source or individualization) is established. However, a sound expression of an opinion about a proposition requires more than a low probability for the evidence alone. Further discussion on this topic is presented Section 2.8. It should be observed that participants of the legal process need to act upon their beliefs, and the question of interest shifts from 'what to believe?' to 'what (or how) to decide?'. These questions are

related as is illustrated by an example presented in Section 2.8.2.

A further problematic statement encountered in the discussions of identification is that the probability of observing another person, in the entire world population, presenting the same analytical characteristics, is zero. Stated otherwise, it is asserted that there is sufficient uniqueness within the observed characteristics to eliminate all other possible donors in the world, and that no contrary evidence (e.g. a solid alibi) can change the expert's certainty. A similar example, this time in ballistics, is the following statement 'numerous courts have prohibited experts from testifying that bullets or cartridge casings were fired from a specific firearm to the exclusion of all other firearms in the world' (*People v. Genrich*, 2019 at p. 58). However, the move from a probability statement to one of certainty represents a 'leap of faith' (Stoney, 1991b) (p. 198), rather than a logical consequence. Stoney (1991b) has famously noted: '[...] are we really trying to prove uniqueness? I would offer to you that it is a ridiculous notion' (p. 198). Assertions of certainty in conclusions thus show a misunderstanding of the role of forensic scientists and of the court in scientific inference procedures, and on the role of statistics in forensic science, see Robertson et al. (2016). For further comments, see also Buckleton (1999), Champod (1999), Taroni and Margot (2000), Buckleton (2004), Stoney (2012), and Biedermann et al. (2013). It is not the role of the

expert to qualify the acceptable level of doubt by supposing a relevant population that, often, is set arbitrarily at its maximum to consider the possibility that all persons, or all firearms, on earth could be the origin of the trace. Participants of the legal process other than scientists should deal with the legal standard, which represents a threshold regarding the issue of identification. In this respect, a discussion on the role of identification in a firearm's cartridge case is presented in *U.S. v. Tibbs* (2019) and *U.S. v. Davis* (2019) where the court concluded by affirming:

> *Experts may not opine that certain cartridge cases were fired by the same gun. (pp. 8–9)*

2.5.10 Other Difficulties

The forensic scientist has to assess the value of the evidence; this means that they have to evaluate the strength of the link between, for example, a recovered trace and a PoI. Therefore, it seems important to point out that forensic evidence evaluation has – by its nature – a close link to statistical assessment (results are associated with probabilities as measures of uncertainty). However, there is a potential for a misinterpretation of the value of statistical evidence when, routinely, such evidence supports a scientific argument in the adversarial (or inquisitorial) system of the trial process. In this section comments are made on the meaning of some concluding statements given by

DNA experts. Note that the probabilistic meaning of these statements is presented assuming that the relative frequency was derived correctly and that there is no possibility of error or misinterpretation of the data. Note also that what follows is part of a survey made in the early 2000s and published in Taroni et al. (2002). Nevertheless, the current situation has not greatly changed (Taroni, 2018) and so those remarks are still of topical interest. More discussion on practical guidance for evidence evaluation can be found in Puch-Solis et al. (2013), Evett et al. (2016), and Gill (2019).

2.5.10.1 Relative Frequency of Occurrence

The use of the relative frequency, γ, is inappropriate when describing a match between two samples of DNA. A typical way of expressing the result, used by several laboratories, is 'The DNA profile in question occurs in about one person in 100 000 of the population'. There are four main objections to this approach:

(1) If the population in question is considerably greater in size than $1/\gamma$ (for example, 3 million), then it might be reasonable for a Court to consider that about $3\,000\,000 \times 1/100\,000$ people in the population would have the same profile. The Court could then perhaps use this DNA evidence to form prior odds when evaluating the remaining evidence in the case. If however the population is much smaller in size than $1/\gamma$ (e.g. a laboratory

reports 1 person in 2.5×10^9), it would not be expected to find anyone else in the population who possesses the profile, and it seems impossible logically to combine DNA evidence with the other evidence in the case which may provide support for the defence hypothesis, for example, convincing evidence of an alibi.

(2) A more serious objection to 'frequency of occurrence' occurs when the scientist is considering the alternative hypothesis that the DNA has originated from a close relative of the PoI. It does not make any sense to say: 'The DNA profile in question occurs in about one brother in 400', so the scientist has to find a different way of expressing this result whilst avoiding confusion for the court.

(3) In considering the numerator and denominator of the likelihood ratio two probabilities are evaluated, the probability of the observation of a match between the crime stain and the PoI, given that the stain has come from the PoI, and given that the stain has come from someone else. Also, consideration of the proposition that the stain has come from someone else is dependent on the result that has been obtained from the PoI. Quotation of a relative frequency may be appropriate when considering a sample of DNA from the scene of a crime with no persons of interest, but is inappropriate when matching DNA has been obtained from a PoI.

(4) There is a range of cases, for example, missing persons, paternity and, in particular, cases where mixed or low quantity DNA profiles have been obtained, where it is not possible to use a relative frequency to express the value of the DNA evidence. The simple reason is that in general the numerator of the likelihood ratio is less than 1, and hence the value of the DNA evidence is not given by $1/\gamma$.

Note that point (3) introduces the notion of the *random match probability* (RMP) or more generally *conditional match probability* (CMP) (Balding and Nichols, 1994), which represents an acceptable way of expressing the value of DNA evidence. It can be used as an alternative to the likelihood ratio in simple cases where corresponding profiles have been obtained, an idea also supported by the ENFSI (2015) guideline that commented:

> *When source level propositions are considered, and when the likelihood ratio amounts to the reciprocal of a conditional match probability - typically in a DNA case involving a large unmixed stain - the forensic practitioner may choose to report the conditional match probability instead of the likelihood ratio. (p. 18)*

For example, the scientist may set out their interpretation as: 'I have considered two propositions with respect to the matching DNA profiles: H_p, the semen has come from the PoI, and H_d, the semen has come from an unknown person who is unrelated to the PoI. These findings are what I would have expected if the first proposition is

true. I have assessed the probability of obtaining corresponding profiles as 1 in a million if the second proposition is true.'

It is also possible to express such a probability without explicitly stating the propositions: 'The genetic profiles of the semen traces are identical to that of the PoI. The probability of an unrelated person of European origin presenting by chance the same genetic profile as the semen stain is about 1 in 1 billion.' or 'The DNA profile of the semen matched that of the blood of the PoI. The chance of a person chosen at random in the population who is not related to the suspect sharing this profile is less than 1 in 10 million.'

Details on the meaning, the calculation, and the magnitude of a random match probability are given in Section 6.1.5 and discussed further in Kaye (1993) and Kaye (2009c).

2.5.10.2 'Could Have' Approach

Several scientists have prefaced an estimate of a relative frequency with phrases like: 'The semen stain could have come from Mister X, the PoI'. 'Sample A could have come from the donor of sample B'. 'The semen stain may originate from the PoI'. 'Based on the results of the DNA analysis it can be concluded that this semen can originate from the PoI'. 'According to the results of the DNA analysis the bloodstain may originate from the victim'. 'According to the results of the DNA analysis the bloodstain could originate from the person in question.'

It may be helpful to the investigator to spell out what may seem obvious, namely, that if the DNA from the crime stain matches the PoI, then the PoI could be the source of the DNA. If this is followed by a statement about the relative frequency of the profile, it is not clear what message is given about the strength of the evidence. A fruitful discussion on this topic is presented in Evett and Weir (1998) and Evett et al. (2000d).

A statement that the evidence 'could have come from the PoI' and similar statements could be seen as a transposed conditional as it is expressing a view about the probability of the proposition. If this type of explanation is considered necessary, then it would be preferable to use a form of words such as 'The DNA profile from the bloodstain matches that obtained from the PoI. Therefore the PoI, or anyone else with the same DNA profile could be the donor of the bloodstain.' This could be read as providing further explanation of the matching profiles, rather than as a probability statement about a proposition.

2.5.10.3 'Cannot Be Excluded' Approach

Among the possible statements, a conclusion that is frequently used is the following: 'The defendant cannot be excluded as the stain donor.' In *U.S. v. Morrow* (2005) it can be read

> *Utilising this analysis, the FBI will conclude that it is scientifically reasonable to attribute the source of a given DNA sample to an individual if the profile frequency of the ostensible source and the matching unknown sample is smaller*

> *than 1 in 280 million.*[10] *If the frequency is higher than this ratio, then the defendant will fall into one of the remaining four categories of results, depending on the exact size of the ratio [the occurrence in a given population], i.e. 'potentially the major contributor in a mixed sample', 'cannot be excluded as a potential contributor of the sample', 'cannot be excluded as a potential major or minor contributor in a mixed sample', and 'excluded as a potential contributor of the sample'. (p. 8)*

Such a statement is close to the previous one ('could have' approach) in its vagueness. But it is also related to a statement typically used in paternity cases where the concept of exclusion is presented in a numerical form using a 'probability of exclusion'. For example, if the characteristic is shared by 0.1% of the population, then the probability of exclusion is 0.999. As clearly explained by Robertson and Vignaux (1992), such a probability tells the scientist what proportion of the population the test would exclude, regardless of who is the father (the donor) of the child (stain). Therefore, this estimate is a measure of the efficacy of the test, because it answers the question 'how likely is this test to exclude Mister X if he is not the father (the donor of the stain)?' However, the court is interested in another question 'how much more likely is the evidence if Mister X is the father (the donor) of the child (stain) than if some randomly selected person were?' The probability of exclusion is not relevant in trying to answer this question.

[10] A comment on this uniqueness aspect has been presented in Section 2.5.9.

Care is required in choosing a form of words which avoids ambiguity. A phrase such as 'The probability of finding another person who has the same genetic profile is 1 in 1 million in the population' could be interpreted as: 'If a DNA profile was obtained from every member of the population, then the probability of finding another person with the same profile is 1 in 1 million'. Clearly if the population is about 50 million, then there is a very high probability of finding someone else with the same profile. A similar example of ambiguous wording is: 'This genotype can be found in 4.07×10^{-10} people in the reference population'.

2.5.10.4 'Consistent with' Approach

Other experts' statements simply note a compatibility between the features observed on the recovered and control materials. One can read sentences like 'Based upon the comparisons the expert performed, they concluded that the hairs found in the victim's apartment were "consistent with" the sample provided by the defendant'.[11] or 'A paediatrician who examined the girls testified that both girls exhibited redness in their labia and vulva regions, which was consistent with sexual abuse.'[12] or 'fibers consistent with a black sweat-shirt owned by the petitioner were found on the victim's bed sheets, matching blue fibres were found on the victim's pink nightgown and on

[11] *People v. Linscott* (1991).
[12] *State v. Hollywood* (1984).

petitioner's blue jeans, microscopically consistent fibres were found on the pink nightgown and on petitioner's underwear'.[13] Formerly in 1991, a court of justice noticed that 'We believe that the simple use of the words "match", "consistent", "could have originated", misrepresented the evidence'.[14]

The problem here is that one has to know whether the scientific observations (the observed features) are also compatible with alternative propositions. Here are some examples: 'An inflamed and irritated vagina and outer vaginal lips in a 4-year-old may be consistent with sexual molestation but it is also consistent with infection,'[15] or 'Other evidence tending to disprove or dispute guilt consists of testimony that the forensic evidence would also be consistent with other theories,'[16] or 'Mr Reisman sets forth two basic propositions in rebuttal to Mr Kirshner, which Defendants seek to exclude. The first proposition is that the forensic evidence is consistent with the scenario of someone opening the file using a different program, copying and pasting the contents into a new word document, and then editing the document to look like the original Hertzig Fax. Mr Reisman's second proposition, which is also necessary if the jury is going to believe this all happened at some time other than

[13] *Holmes v. South Carolina* (2006).
[14] *People v. Giangrande* (1991).
[15] *In re Michelle I.* (1993).
[16] *Swearingen v. State* (2003).

the middle of the night, is that system clocks were sometimes unreliable, so that a timestamp of 4:41 a.m. might be inaccurate.'[17]

A detailed analysis of reports and oral testimonies expressing the 'Consistent with' approach is made by Edmond (2013). A comment and a solution is presented by Lyon and Koehler (1996). The authors wrote:

> *A relevance ratio analysis, however, reveals that a condition that is 'consistent with' abuse is relevant for proving that the abuse occurred only when the condition occurs more frequently among abused children than among non-abused children. Typically, the 'consistent with' terminology is merely an observation that at least some abused children exhibit the condition. Thus, 'consistent with' testimony informs a factfinder that the numerator of the relevance ratio is nonzero, but says nothing about the denominator. The numerator must be compared with the denominator, however, for the relevance of the condition to be fully understood. (at p. 51)*

Note also that Lyon and Koehler (1996) affirmed that 'in [their] view, a relevance ratio analysis is the most efficient way to think about evidentiary relevance.' (p. 47). They wrote:

> *Readers familiar with the literature on probabilistic reasoning in the law may recognise the relevance ratio as the likelihood ratio term from Bayes' theorem. Bayes' theorem provides a method for updating probabilistic beliefs in the face of new evidence. It combines a likelihood ratio (which*

[17] *Robocast, Inc. v. Microsoft Corp.* (2014).

> *captures the diagnostic value of new evidence) with a prior odds ratio (which captures one's initial beliefs about the hypothesis) to form a posterior odds ratio (post-evidentiary belief about the hypothesis). (p. 48)*[18]

This wording refers to what has been presented in Section 2.3.

Further comment came from Evett et al. (2000d) wrote:

> *The phrase 'consistent with', when used in the context of 'the evidence is consistent with this shoe having left the mark' is a potential source of confusion. Also it is not balanced, unless the other proposition is addressed. Perhaps it might be equally true to say 'the evidence is also consistent with any other shoe of the same model and size having left the mark'. At worst, the use of this phrase lays one open to the criticism of partiality. At best, it does nothing to convey an assessment of the weight of evidence in favour of one or other of the stated propositions. (p. 237)*

The term 'consistent with' can represent an adequate way to express the evidence if the expression is also associated to at least one alternative proposition. An example is the following:

> *Based on Dr. Wolf's testimony and the amended autopsy report, the State stipulated that the credible forensic evidence was 'more consistent with the theory that Pena was shot at the location where he was found,' as opposed to some other location.*[19]

[18] The authors meant 'odds' here not 'odds ratio'.
[19] *Ex parte De La Cruz* (2015).

2.5.11 Empirical Evidence of Errors in Interpretation

Scientific evidence is often presented in a numerical way. Such an evaluation inevitably uses probabilities as measures of uncertainty. Judges are concerned that scientific evidence may overwhelm or mislead the jury, especially when its presentation by an expert may appear to give the evidence greater probative value to a layman than it would to another expert (see, for sake of illustration the analysis presented by Koehler (2014)). This concerns the legal community when a decision rests on a resolution of the differences between opposing experts, with seemingly conflicting testimony. The potential for basing a decision on a misunderstanding is considerable when the uncertainty of the scientific evidence is not understood (see, for example, (Fienberg et al., 1996); (Kaye and Koehler, 2003)). Statistical forms of scientific evidence have probably yielded the greatest confusion and concern for the courts essentially in the application of DNA (Kaye, 1993; National Research Council, 1996). The confusion is not surprising; the courts still have little expertise in genetics and statistics. This is the major reason for the need to publish specific guidelines for evidence evaluation as proposed by the Royal Statistical Society (see Puch-Solis et al. (2013)) and the Royal Societies of London and

Edinburgh[20] (Royal Society and Royal Society of Edinburgh, 2017) and still supported by scientists faced with practice and consultancy (Evett et al., 2016). Koehler (2018) summarised the current flawed situation and provided trial judges with guidance on how they should think about and evaluate the reliability of forensic science evidence.

Scientists have also provided sources of misinterpretation in their reports and statements (Koehler, 1993a). Moreover, it has also been argued that the presentation at trial of the evidential value in the form of a likelihood ratio could be very prejudicial in the decision-making process (Koehler, 1996). Koehler noted that '(e)ven when likelihood ratios are properly conveyed, there is little reason to believe that jurors will understand what they mean and how they should be used. Although they have scientific merit, likelihood ratios – which are the ratios of conditional probabilities – are not easy to understand.' Psychological research has emphasised the fallacious way in which people reason in managing uncertainty and probabilities, especially with conditional probabilities (for a review of these studies see Kaye and Koehler (1991), Fienberg and Finkelstein (1996), Edmond et al. (2017) and Martire and Kemp (2018)).

Empirical research has been carried out in the last decade. Following from the results, methods of improving the reporting of the evidence and the

[20]http://www.rss.org.uk/statsandlaw.

trial presentation have been proposed in order to assist with the questions

- Is the evidence correctly interpreted?
- Can the way in which the evidence is presented at trial influence a verdict?
- Can the way in which the evidence is presented at trial influence an update of a probability of guilt?
- Can an interpretation of the value of the evidence be misunderstood?

Cases have been studied where scientific evidence has been presented in court, and, using information gained from such studies, a series of problems have been developed for, and given to, law and forensic science students, to practitioners (advocates and forensic scientists,) and to mock jurors to investigate their understanding of uncertainty. Problems associated with the presentation of scientific evidence at trial were investigated using the responses of students and practitioners. Research studied the interpretation of numbers related to the value of scientific evidence, numbers that had been used by experts to explain the value of the evidence. For example, the impact of different ways of presentation and the value of similarities between a DNA recovered trace and a DNA control material on the verdict (guilty or not guilty) and on the update of the probability of guilt were studied. The results showed an underestimate of the value as expressed by the

posterior probabilities computed using Bayes' theorem. The fact that these posterior assessments were substantially below those computed from Bayes' theorem confirmed results of previous studies (details in Taroni and Aitken (1998b), Koehler (1996)). For a review of earlier studies, see Fienberg and Finkelstein (1996). The results also suggested that subjects did not treat different methods of presentation of the evidence (percentage of exclusion, relative frequency, likelihood ratio and posterior odds) similarly but that there was an association between the assessment of the posterior probabilities and the verdicts. Moreover, subjects did not seem capable of distinguishing between the magnitude of the difference in values between scenarios and the effect of error rate, if reported (Koehler et al., 1995; Koehler, 2001a). Even if the results have to be treated with caution, because of the limits of sample size and geographical areas studied, they show a clear problem in dealing with measures of uncertainty (Koehler, 2001b). Studies are described in which (Koehler, 2001b)

> *DNA match statistics that target the individual suspect and that are presented as probabilities (i.e. 'the probability that the suspect would match the blood drops if he were not their source is 0.1%') are more persuasive than mathematically equivalent presentations that target a broader reference group and that are framed as frequencies (i.e., 'one in one thousand people in Houston would also match the blood drops'). (p. 493)*

Other empirical research has approached the problem of the possible pitfalls of intuition related to the presentation of scientific evidence in numerical ways. The available DNA data (as collected in population studies) allow the scientist to offer to the court a number that should quantify the strength of the link generally established between a PoI and a trace recovered on a victim or on a crime scene. This number generally represents the relative frequency of the matching characteristic in some relevant population (or the RMP, Section 6.1.5). The use of these numbers may be thought prejudicial. Conclusions based solely on the relative frequency of the matching trait could have serious consequences as already discussed in Sections 2.5.1–2.5.6 where the scientists' statements use subtle forms of reasoning. Research has tried to measure the extent of the misunderstanding of the meaning of the value of the statistical evidence. Generally, the experts' statements in criminal trials were presented to participants in a survey. Experts gave different explanations of the meaning of the statistical evidence presented. Participants were asked to detect which statements were correct and which ones were erroneous. Where the expert's explanation was thought to be erroneous, participants were asked to explain their reasons. A correct answer was taken to be one in which the respondent says the expert is correct when the expert is indeed correct or the respondent says the expert is wrong when indeed the expert is wrong (all details of

cases and analyses can be found in Taroni and Aitken (1998b–1999). The concern here is with principles. For example, in *E. Ross v. State of Indiana* (1996) (Section 2.5.2), *R v. Michael Gordon* (1994), and *R v. Deen* (1993), the problem was the fact that the expert gave an opinion on an issue (the PoI is the source of some evidence, or the PoI is the rapist). These are examples of the transposed conditional. Participants in the surveys were confused by the statements and unfortunately accepted them. In *R v. Montella* (1992), the expert presented the value of the evidence as a likelihood ratio. This logical method of assessing and presenting the evidence unfortunately created considerable confusion in the comprehension of the expert's statement. There were comments from members of all groups of participants (students and practitioners) that they believed the explanation to be wrong, confusing, and too difficult to understand. The results were supported by Koehler (2001b). The aim of this book is to reduce, or even eliminate, the incidence of these problems. In *U.S. v. Jakobetz* (1992), the expert fell into a *source probability error* (Section 2.5.2). This equates the frequency of the trait with the probability that someone other than the defendant was the source of the trace evidence. The case *Ross v. State* (1992) presents another statistical problem. This fallacy is an example of the *numerical conversion error* (Section 2.5.6) because it may be thought that the significance of the value of the relative frequency can be measured

by equating the reciprocal of the frequency with the number of people who would have to be tested before there is another match. Generally, results showed that participants believed that experts were right in their explanation of the evidence. In this last situation, the participating members of the Faculty of Advocates of Scotland correctly identified the statistical meaning of the evidence presented stating that 'just because the odds (*sic*!) are 1 to 209 100 000 it doesn't mean that you have to look at 209 100 000 people before finding a match – the next sample could be the next match'. From a research point of view, studies with mock jurors and/or students have been essentially focused on management of two distinct pitfalls of intuition: the *prosecutor's fallacy* and the *defence attorney's fallacy* (Sections 2.5.1 and 2.5.4, Thompson and Schumann (1987), Thompson (1989), Carracedo et al. (1996)). These studies involved mock trial scripts rather than actual trials.

Other research submitted real criminal cases, where the statistical evidence had been explained, to specialist groups such as students (who represent future judges, lawyers and forensic scientists) and as practitioners (forensic scientists and advocates). Results in both kinds of research showed that the great majority of participants failed to detect the error in the arguments exposed by experts at trial. The tendency to draw erroneous conclusions from fallacious descriptions of the meaning of the evidence is

troubling. It demonstrates a lack of knowledge in conditional probability, knowledge that is required to assess correctly the value of the evidence and to appreciate correctly the meaning of this value.

Consider fibres and glass evidence. As empirical research (Champod and Taroni, 2017) has supported, risks of misconception exist when experts use relative frequencies to support their analytical results. In fact, the relative frequency is only one of many parameters that should be considered in a complete perspective of evidence evaluation (see Chapter 5).

The use of the likelihood ratio constitutes, for the expert, an interesting subject for reflection on scientific proof, because the expert must search and choose the relevant questions considering the physical evidence from two opposite points of view. Case surveys represent an attempt to study the evaluation framework for different scenarios involving fibres, blood, and glass fragments (Taroni and Aitken, 1998c). These scenarios were chosen deliberately in order to illustrate the assessments and the evolution of the different parameters in different situations. Subjects' reactions to a situation in which more than one foreign group of fibres was recovered at a scene of crime and determined to have been left by the offenders were studied. Only one of them (group of fibres) was compatible with an article of clothing associated with a PoI. Evaluation of the match between recovered fibres and clothing from a PoI must consider similar and dissimilar elements as

already suggested for bloodstain evidence (Evett, 1987b). Therefore, it is important not only to focus on the fibres that match the garments of the PoI, but also to consider other groups of fibres compatible with the facts, which potentially could have been left by the offender, who is not the PoI; see also Section 6.1.2 for a discussion.

Two scenarios with different values for the fibre evidence were described to participants who were asked for assessments (Evett, 1983; Buckleton and Evett, 1989). It was found that subjects did not change their assessment of the evidence according to the difference between the scenarios. Subjects did not take into account, in their assessment of the value of the evidence, the number of groups of fibres that were compatible with the issue in question. These failings produced an overestimate of the value of the evidence in the case where there was more than one distinct group of fibres (details of the likelihood ratio development are in Sections 2.3 and 2.4). When two individuals (or an individual and an object) have contact during a criminal action, a reciprocal transfer of material (e.g. fibres or glass) is involved. Where this happens, the two sets of recovered traces have to be considered as dependent. If a transfer has occurred in one direction, and the expert has recovered traces characterising this transfer, then the expert would expect to find trace evidence characterising the transfer in the other direction. The presence of one set of transfer evidence gives information

about the probability of the presence of the other set of transfer evidence (Inman and Rudin, 2001) (details of such a situation are presented in Section 5.3.2.5). The ability of participants to distinguish the scenario where the two sets of recovered traces are dependent from the scenario where the two sets of recovered traces are independent was investigated, with particular reference to the reaction to new technical information about the presence or the absence of cross-evidence involved in the criminal contact. The absence of the expected presence of some trace evidence which would show a reciprocal exchange of material has to be taken into account in the assessment of the value of the recovered matching evidence. The results obtained in the surveys supported previous results and emphasised the subjects' inability to take into account technical information in the assessment of the real value of a link detected between two persons or objects (Taroni and Aitken, 1998b). The assessment by the participants of the probative force of an entire aggregation of evidence was also investigated. In one scenario, participants were asked to make an aggregated judgement, which concerned a large collection of evidence in a criminal case involving glass evidence presented by the experts for the prosecution and for the defence. In a second scenario, for comparison with the first, the participants made assessments for two subsets of the evidence and these were combined to provide an overall judgement. This

required more assessments, but each one was made with reference to a smaller and more specific body of evidence. A comparison was made of the assessments of posterior probabilities made by the participants in each of the two scenarios. The posterior probability of guilt made by participants who received the entire body of evidence was much smaller than that made by the participants who updated their probability twice. The results indicate that arguments tend to provide smaller posterior probability assessments if the body of evidence is not decomposed. Note also that, in both scenarios, smaller values for the probability of guilt were obtained than would have been if the laws of probability had been followed.

In general, results have shown that methods of assessment used by participants are insufficient to obtain a correct value of the scientific evidence. Dangers of underestimation and overestimation of the real value of the evidence still remain. Forensic scientists endeavour to give the court an accurate evaluation to illustrate the true worth of their results. Unfortunately, judging from the results of surveys, the evaluations of the scientists fail to consider all the parameters involved in the scenarios proposed. Furthermore, comments on the calculation of the posterior probabilities show that these have been based upon 'a subjective decision' instead of in accordance with the rules of probability. As stated earlier, the studies were designed to answer four questions

- Is the evidence correctly interpreted?
- Can the way in which the evidence is presented at trial influence a verdict?
- Can the way in which the evidence is presented at trial influence an update of a new probability of guilt?
- Can the explanation of the evidence be misunderstood?

The studies were pertinent as the National Research Council of the United States had commented that 'there is a lack of research into how jurors react to different ways of presenting statistical information' and that 'no research has as yet tested the reactions of triers of fact to the detailed presentations of evidence on DNA profiling that are encountered in the courtroom' (National Research Council, 1996). On the same line of reasoning, see the PCAST Report (President's Council of Advisors on Science and Technology (PCAST), 2016).

The answers to the questions were all undesirable: the evidence may not be correctly interpreted, the way in which the evidence is presented at trial may influence a verdict, the way in which the evidence is presented at trial may influence an update of a new probability of guilt and the explanation of the evidence may be misunderstood. This book presents the view that the likelihood ratio (in the Bayesian framework) should be used by experts because it allows them

to take into consideration the evidence under two alternative propositions and it enables the consideration of other relevant factors in the calculation of the value of the evidence (as will be presented in the following chapters). Jurists should also appreciate the approach, because it clarifies the roles of the expert and of the judge or members of the jury: the latter take the decision on an issue, the former compares the likelihoods of the evidence under two proposed propositions.

It is important to realise that in the evaluation of evidence the probability of the evidence has to be considered under two propositions, separately. The following quotes illustrate this point (Friedman, 1996):

> *The concept of a match is gratuitous.*
>
> *The factfinder's task is to assess the relative probability of two hypotheses – that the samples came from a common source, and that they did not.*
>
> *The evidence is the two profiles revealed by the samples.*[21]

[21] It is of interest to note the conclusions expressed by some judges concerning the use of the term *match*. For example, in U.S. v. Davis (2019) at p. 7 it can be read: 'Concerns over the reliability of this testimony expressed in the National Research Council (1996) and President's Council of Advisors on Science and Technology (PCAST) (2016) reports and those reflected in a recent chorus of federal decisions lead the court to impose certain restrictions on the testimony of these toolmark examiners. The examiners may not testify that the marks indicate a match or that cartridge cases were fired by the same firearm. They may not testify that cartridge cases have signature toolmarks identifying a single firearm.' In the same case, at p. 9, the judge wrote: 'Experts may not opine that a cartridge case is a *match* to other cartridge cases or firearms.'

> *A factfinder can ask how likely it is that the evidence would have arisen, given each of the competing hypotheses, without asking whether the evidence satisfies an arbitrary defined match standard. (p. 1826)*

and

> *This concern might have some theoretical force, in the context of DNA evidence, when a prosecutor presents evidence that two samples do not match because the disparity between the measurements is so great. Such a conclusion tells the factfinder that the evidence would be unlikely to arise given the hypothesis that the samples had a common origin, but it does not combine easily with other evidence because it does not tell the factfinder how likely the evidence would arise given an alternative hypothesis. (p. 1827)*

2.6 MISINTERPRETATIONS

The aforementioned examples are appropriate summaries of the evidence. However, misinterpretations still occur in which the evidence is summarised as a comment about the truth or otherwise of the prosecution's proposition. Note that the use of verbal scales about the truth or otherwise of a proposition are accepted in many scientific fields (e.g. medicine, weather forecasting). Here, in contrast to forensic science, the scientist plays a different role and uses a different amount of information. Therefore, posterior scales in these disciplines seem acceptable in a way they are not in forensic science. An 11-point subjective posterior scale of scientific uncertainty based

on legally defined standards of proof has been proposed (Weiss, 2003).

In a response to a survey conducted by Taroni and Aitken (2000) on fibres evidence, the comment was made that the strength of the evidence was categorised in terms of the probability of the prosecution's proposition, that it was

- Beyond reasonable doubt,
- Most probable,
- Probable,
- Quite possible or
- Possible

that matching evidence associated with the defendant comes from the same source as that found at the crime scene. In this survey, laboratories generally commented on the truthfulness or otherwise of the proposition proposed by the prosecution. This was instead of the value of the evidence.

Also, in the context of human hair comparisons, Gaudette (2000) gave a scale for the questioned hairs originating or not from the same person as the known sample. There is a match and the scale interprets this as

Strong positive:	Questioned hairs originated from same person as the known sample.
Normal positive:	Questioned hairs are consistent with the known sample.

Inconclusive:	No conclusion can be given.
Normal negative:	Questioned hairs are not consistent with the known sample.
Strong negative:	Questioned hairs could not have originated from the known sample.

However, this is making a judgement about the source of the hair without prior knowledge of the background information of the case. Similar problems have been discussed in the context of shoeprint examinations (Champod et al., 2000; Taroni and Margot, 2001; Katterwe, 2002a,b; Taroni and Buckleton, 2002; Champod and Jackson, 2002) and speaker recognition (Champod and Evett, 2000) and are still under discussion in some areas, see, for example, Morrison et al. (2016) where the authors also presented the use of different frameworks for reporting the conclusions of speaker identification analyses.

Finally, the difficulties associated with probabilistic reasoning in the criminal justice system are illustrated by this contradiction within the same judgement. The court in *R. v. France* (2019) ruled by affirming that the expert

> *. . . would not be permitted to testify as to whether an assault was more likely to have caused the injury than an accidental fall, nor would he be permitted to express an opinion on the probabilities of one cause as opposed to another. (p. 2)*

This is sound advice as discussed in Section 2.5.1. Unfortunately, one page later, the court ruled that

> *[the expert] should avoid using language to say that the injuries were 'consistent with' an assault, but rather should use language such as the injuries could have been caused this way, or it is possible they were caused this way. (p. 3)*

2.7 EXPLANATION OF TRANSPOSED CONDITIONAL, DEFENCE ATTORNEY'S AND FALSE POSITIVE FALLACIES

Using the odds form of Bayes' theorem (2.14) insight into the fallacy of the transposed conditional (or prosecutor's fallacy), the false positive fallacy and the defence attorney's fallacy can be gained.

2.7.1 Explanation of the Fallacy of the Transposed Conditional

As before, a crime has been committed. A stain of blood has been found at the scene, which has been identified as coming from the criminal. A PoI is identified and their blood group is the same as that of the crime stain. Let E denote the evidence that the blood group of the PoI is the same as that

of the crime stain. Let H_p denote the proposition that the PoI is guilty and its complement H_d denote the proposition that the PoI is innocent.

Consider the following two statements:

- The blood group is found in only 1% of the population,
- There is a 99% chance that the suspect is guilty.

The second statement does not follow from the first without an unwarranted assumption about the prior odds in favour of guilt. The first statement that the blood group is found in only 1% of the population is taken to mean that the probability that a person selected at random from the population has the same blood group as the crime stain is 0.01. Thus $\Pr(E \mid H_d) = 0.01$. Also, $\Pr(E \mid H_p) = 1$, assuming no false negatives. The value of V is 100.

The second statement is taken to mean that the posterior probability in favour of guilt (posterior to the presentation of E) is 0.99; i.e. $\Pr(H_p \mid E) = 0.99$. Thus $\Pr(H_d \mid E) = 0.01$ since H_p and H_d are complementary propositions. The posterior odds are then 0.99/0.01 or 99, which is approximately equal to 100. However, V also equals 100. From (2.14), the prior odds are approximately equal to 1:

$$\Pr(H_p) \simeq \Pr(H_d).$$

The second statement of the fallacy of the transposed conditional follows from the first only if

$\Pr(H_p) \simeq \Pr(H_d)$. In other words, the PoI is just as likely to be guilty as innocent. This is not in accord with the dictum that a person is innocent until proven guilty. The prosecutor's conclusion, therefore, does not follow from the first statement unless this unwarranted assumption is made.

The use of prior odds of 1 was also advocated for shoe print examination (Katterwe, 2003). A population of N shoes is postulated as the one to which the shoe that made the print at a crime scene belongs. There is one shoe that may be considered as the suspect shoe. In the absence of any other information, the probability that this shoe made the crime print is $1/N$. The probability that another shoe from this population made the print is $(N-1)/N$. The prior odds that the suspect shoe made the print are $1/(N-1)$. The argument advanced by Katterwe (2003), however, is that, apart from the suspect shoe and in the absence of any other information and with the assumption that shoeprints are unique, there is a population of only one shoe that could also have made the print. The defence proposition is that only one other shoe made the print. Therefore the relevant population is of size two. This argument is analogous to the probability of hitting a target, for example, in a game of darts. A dart may hit the bulls-eye or it may not. There are only two possibilities. In the absence of other information, the probabilities of 'hit' or 'miss' are equal at $1/2$ each. However, in reality, there is always other information. For the dart player, there is information about the area of

the bulls-eye compared with the area of the rest of dart-board and wall on which the board hangs. For shoe-print examination there is information on the number of shoes in the world that could have made the print. Similar arguments hold for other evidence types, e.g. paternity, Section 6.3.4.

2.7.2 Explanation of the Defence Attorney's Fallacy

Assume the likelihood ratio is 100, as in the discussion of the fallacy of the transposed conditional. Consider a relevant population to contain 200 000 people. The defence says that there are 2 000 people with the same blood group as the defendant. The probability the defendant is guilty is 1/2 000 and thus the evidence has very little value in showing this particular person guilty. As before $\Pr(E \mid H_p) = 1, \Pr(E \mid H_d) = 0.01$ and V equals 100. Also, $\Pr(H_p \mid E) = 1/2\,000$ and so $\Pr(H_d \mid E) = 1\,999/2\,000$. The posterior odds in favour of H_p are

$$\frac{1/2\,000}{1\,999/2\,000} = \frac{1}{1\,999} \simeq \frac{1}{2\,000}.$$

The prior odds equals the ratio of the posterior odds to V:

$$\begin{aligned}\frac{\Pr(H_p)}{\Pr(H_d)} &= \frac{\Pr(H_p \mid E)}{\Pr(H_d \mid E)}/V \simeq \left(\frac{1}{2\,000}\right)/100 \\ &= \frac{1}{200\,000}.\end{aligned}$$

Thus, $\Pr(H_p) = 1/200\,001$, $\Pr(H_p) = 200\,000/200\,001$. The prior probability of guilt is $1/200\,001$. The denominator is the size of the relevant population (of innocent people) plus one for the criminal. The implication is that everybody is equally likely to be guilty. This does seem in accord with the dictum of innocent until proven guilty. Colloquially, it could be said that the defendant is just as likely to be guilty as anyone else. The defence attorney's fallacy is not really a fallacy. It is misleading, though, to claim that the evidence has little relevance for proving the suspect is guilty. Evidence that increases the odds in favour of guilt from $1/200\,000$ to $1/2\,000$ is surely relevant. Using the scale presented in Table 2.8, the likelihood ratio expresses a moderate support to the proposition of guilt.

Notice that it is logically impossible to equate the dictum *innocent until proven guilty* with a prior probability of guilt of zero, $\Pr(H_p) = 0$. If $\Pr(H_p) = 0$, then $\Pr(H_p \mid E) = 0$, from (2.11) no matter how overwhelming the evidence, no matter how large the value of V. The probability of guilt may be exceedingly small, so long as it is not zero. So long as the prior probability of guilt is greater than zero, it will be possible, given sufficiently strong evidence, to produce a posterior probability of guilt sufficiently large to secure a conviction. If the defendant is as likely to be as guilty as anyone else; his prior probability of guilt would then be the reciprocal of the size of the population defined by 'anyone else'. Comments on

this point are presented in Robertson and Vignaux (1994). Lindley (1985) presented this aspect as Cromwell's rule[22] by stating:

> *A simple result that follows from Bayes' theorem is that it is inadvisable to attach probabilities of zero to uncertain events, for if the prior probability is zero so is the posterior, whatever be the data. This is immediate since the latter is proportional to the product of the likelihood with the former, and a product is necessarily zero if one of its factors is. Consequently an uncertain event of zero probability remains so whatever information is provided. In other words, if a decision-maker thinks something cannot be true and interprets this to mean it has zero probability, he will never be influenced by* any *data, which is surely absurd. (p. 104)*

2.7.3 Explanation of the False Positive Fallacy

It is important to have accurate information about both the RMP (see Section 6.1.5 for details) and the false positive probability (see Section 6.1.6.4 for details) when evaluating evidence. Ignorance of, or an underestimation of the potential for, a false positive can lead to serious errors of interpretation, particularly when the other evidence against the PoI is weak.

It is considered essential to have valid scientific data for determination of a RMP but, paradoxically, it is thought unnecessary to have valid data

[22]This is called Cromwell's rule because of his advice to the Church of Scotland 'I beseech you … think it possible you may be mistaken' (Cromwell, 1979).

for determination of a false positive probability. The explanation for this lies partly in the *false positive fallacy* (Section 2.5.7). It is assumed, mistakenly, that if the false positive probability is low then the probability of a false match must also be low in every case. For example, a forensic scientist who thinks that there is only a 1% chance of falsely declaring a match between samples in a case if they really do not match, might assume that there is, necessarily, a 99% chance that the reported match (RM) is a true match. This assumption is fallacious. The fallacy arises from a mistaken equation of the conditional probability of a match being reported when the samples do not match (the false positive probability) with the probability that the samples do not match when a match has been reported. These two probabilities are not the same. The false positive probability is the probability of a match being reported under a specified condition (no match). It does not depend on the probability of the occurrence of that condition. By contrast, the probability that the samples do not match when a match has been reported depends on both the probability of a match being reported under the specified condition (no match), and on the prior probability that condition will occur. Consequently, the probability that a reported match is a true match or a false match cannot be determined from the false positive probability

alone. In formal terms, the fallacious assumption is that $\Pr(M \mid R) = 1 - \Pr(R \mid \bar{M})$, where M is the event that the PoI and the perpetrator have matching DNA profiles, $\bar{M}$ is the event that they do not have matching profiles, and $\Pr(R \mid \bar{M})$ is the false positive probability; i.e. the probability of a match being reported given that the samples do not have matching profiles. This assumption is fallacious because it ignores the prior odds that the PoI's profile matches the sample profile. Let the prior odds, $\Pr(M)/\Pr(\bar{M})$, equal $1/k$ where k is large. Then

$$\frac{\Pr(M \mid R)}{\Pr(\bar{M} \mid R)} = \frac{\Pr(R \mid M)}{\Pr(R \mid \bar{M})} \times \frac{1}{k}.$$

Assume $\Pr(R \mid M) = 1$; i.e. there are no false negatives. Then

$$\Pr(M \mid R) = 1/\{1 + k \Pr(R \mid \bar{M}\}$$

which can be much lower than $1 - \Pr(R \mid \bar{M})$ when k is large.

For example, suppose that the prior odds the PoI's profile will match that of the recovered stain are $1/1\,000$ because the PoI is selected through a large DNA dragnet and appears, initially, to be an unlikely perpetrator. Suppose further that a DNA match is reported and that the false positive probability is 0.01. The probability that this reported match is a true match is, therefore,

$1/(1 + 1\,000 \times 0.01) = 0.0999$. In other words, the probability that this reported match is a true match is not 0.99, as the false positive fallacy would suggest; it is less than 0.1.

True matches are expected to be rare when a database is searched. Therefore, the probability in a particular case that a non-match will mistakenly be reported as a match, even if low, may approach or even surpass the probability that the suspect truly matches. The false positive fallacy is similar in form to the prosecutor's fallacy (Thompson and Schumann, 1987), but differs somewhat in content (Thompson et al., 2003).

Victims of the false positive fallacy mistakenly assume that $\Pr(M \mid R) = 1 - \Pr(R \mid \bar{M})$. Victims of the prosecutor's fallacy mistakenly assume that $\Pr(S \mid M) = 1 - \Pr(M \mid \bar{S})$ where

- S is the proposition that the specimen came from the PoI and
- $\bar{S}$ is the proposition that the specimen did not come from the PoI

(Thompson and Schumann, 1987). Both fallacies arise from failure to take account of prior probabilities (or odds) when evaluating new evidence; both can lead to significant overestimation of the posterior probability when the prior probability is low. The prosecutor's fallacy is an erroneous way of estimating the probability that the PoI is the source of a sample based on evidence of a matching characteristic; the false positive fallacy

is an erroneous way of estimating the probability of a true match based on a reported match.

2.8 MAKING COHERENT DECISIONS

Quantification of uncertainty represents a fundamental step in any forensic case and probabilities represent the rational way to handle it, thus helping to avoid pitfalls of intuition. However, this rarely represents the end of the matter on which a forensic scientist may be asked to work (Lindley, 1977a). There may be cases of interest where the expert needs to make a choice among alternative courses of action (e.g. processing or not a fingermark, pronouncing the exclusion of an individual as the donor of a stain based on the inspection of genetic characteristics). Further, a difficulty that must be tackled is that consequences, unknown at the time of decision-making, might arise from alternative courses of action (e.g. a false exclusion). Decision theory is a general framework wherein both the management of uncertainty with the use of probability and decision-making can be formally analysed. Kingston and Kirk (1964) were aware of this procedure. They wrote:

> *A decision is made by utilising this probability [defined as a belief], in conjunction with considerations as to the consequences of the decision, as a guide. (p. 515)*

2.8.1 Elements of Statistical Decision Theory

The Bayesian paradigm (see Section 2.2) can be extended to encompass the wider perspective of decision making. A decision problem exists whenever there are two or more possible decisions, and there is uncertainty regarding the consequences that may arise from each decision. The principal issue is the selection of a decision. The basic elements of a decision problem are given by the following:

(1) A set of exhaustive and mutually exclusive decisions (or *courses of action*[23]).

Decisions can be denoted d, whilst the set of all decisions (the *decision space*) is denoted $\mathcal{D}$.

(2) A set of exhaustive and mutually exclusive uncertain events (usually called states of nature). States of nature are denoted θ, whilst the set of all possible states of nature (the *parameter space*) is denoted Θ.

(3) A set of consequences, defined as the outcome following the combination of decision d when the actual state of nature is θ. Consequences are denoted $c(d, \theta)$, or simply c, whilst the set of all consequences is denoted $\mathcal{C}$.

See Lindley (1985) for a broad description of the general structure of a decision problem.

[23]Berger (1985) noticed that 'Decisions are more commonly called *actions* in the literature' (p. 3).

The main purpose of a theory of decision making is to conceive of a normative framework that allows decision makers to assess the consequences of alternative courses of action, compare them and avoid irrational choices or behaviour. Principles that should be followed by a decision maker who aims to achieve coherent decision making can be summarised as follows (see e.g. Bernardo and Smith (2000), DeGroot (1970)):

(1) The uncertainty about states of nature θ should be expressed in terms of a probability distribution describing their plausibility, $\Pr(\theta \mid I)$, where I denotes the relevant information available at the time when the probability assessment is made;

(2) The decision maker can express preferences among possible consequences, say, c_1 and c_2, meaning that they must be able to specify at any point that one is preferred or whether they are equivalent. This means that for any pair of consequences $(c_1, c_2) \in \mathcal{C}$, it is assumed a decision maker can always say whether they are indifferent among them ($c_1 \sim c_2$), whether one is strictly preferred to another (e.g. $c_1 \prec c_2$ means that c_2 is strictly preferred to c_1), or whether one is not preferred to another (e.g. $c_1 \preceq c_2$ means that c_1 is not preferred to c_2, that is either $c_1 \sim c_2$ or $c_1 \prec c_2$ holds).

(3) Preferences among consequences should be measured by a *utility function* $U(\cdot)$, which specifies, on some numerical scale, their

desirability. In other words, if $U(\cdot)$ is a utility function and $c_1 \preceq c_2$, then $U(c_1) \leq U(c_2)$. Note that the existence of such a function is conditioned on the acceptance of a set of conditions (axioms) characterising the preference system (DeGroot, 1970).

(4) The desirability of alternative courses of action is measured by their corresponding expected utility, which is obtained by combining utilities $U(c(d, \theta))$ associated with the consequences of decisions $c(d, \theta)$ and probabilities for states of nature $\Pr(\theta \mid I)$

$$\mathrm{EU}(d) = \sum_{\Theta} U(c(d, \theta)) \Pr(\theta \mid I). \qquad (2.18)$$

A standard decision rule instructs one to select the action that maximises the expected utility. This decision is optimal because it can be proved that it is the decision, which has associated with it the highest probability of obtaining the most favourable consequence (Lindley, 1985).

It is often convenient to express preferences among decision consequences $c(d, \theta)$ in terms of a non-negative *loss function* $L(\cdot)$ defined by

$$L(c(d, \theta)) = \max_{d \in D} U(c(d, \theta)) - U(c(d, \theta)).$$

The loss $L(c(d, \theta))$ for a given consequence $c(d, \theta)$ is defined as the difference between the utility of

the best consequence under the state of nature at hand and the utility for the consequence of interest. Stated otherwise, the loss measures the penalty for choosing a non-optimal decision, also called *opportunity loss* (Press, 1989): that is, the difference between the utility of the best consequence that could have been obtained and the utility of the actual one. The loss $L(c(d, \theta))$ thus measures the undesirability of decision d when θ turns out to be the true state of nature. Therefore, the undesirability of alternative courses of action will be measured by their corresponding *expected losses*:

$$\mathrm{EL}(d) = \sum_{\Theta} L(c(d, \theta)) \Pr(\theta \mid I). \tag{2.19}$$

The optimal decision, the one that maximises the expected utility, may be thought of as the one that minimises the expected loss (Savage, 1954). Information I is often omitted to simplify notation, though it is important to keep in mind that it conditions all probability assignments.

2.8.2 Decision Analysis: An Example

Consider the following forensic example presented in Bozza et al. (2014) regarding the chemotype (the chemical characteristics of a substance) of cannabis plants among seized materials. The two propositions of interest are:

H_1: the seized plant is of drug type (population 1);

H_2: the seized plant is of fibre type (population 2).

In a Bayesian perspective, let π_1 denote the prior probability of proposition H_1, $\pi_1 = \Pr(H_1)$, and let π_2 denote the prior probability of proposition H_2, $\pi_2 = \Pr(H_2)$. These two probabilities express the uncertainty about whether the seized material is of type 'drug' or 'fibre', respectively, given the circumstantial information I (omitted from notation for simplicity). As noted in Section 2.1.3, the ratio π_1/π_2 of the prior probabilities of propositions H_1 and H_2 is called the prior odds of H_1 to H_2. The prior odds indicate whether *a priori* proposition H_1 is more or less probable than proposition H_2 (prior odds being larger or smaller than 1), or whether the two propositions are almost equally probable (prior odds close to 1). Following the receipt of evidence E, the posterior probability of proposition H_1, $\Pr(H_1 \mid E)$, denoted α_1 and the posterior probability $\Pr(H_2 \mid E)$ of proposition H_2, denoted α_2, can be computed according to Bayes' theorem as shown in Section 2.2.2.

Note also that the Bayes' factor is defined as the ratio of the posterior odds, α_1/α_2, to the prior odds, π_1/π_2. It measures the change produced by the data in the odds when going from the prior distribution to the posterior distribution.

In a Bayesian decision-theoretic perspective, let $\mathcal{D} = \{d_1, d_2\}$ denote the decision space, where $d_{1(2)}$ represents the decision of classifying the plant

material available for examination in population 1(2). Decision $d_{1(2)}$ is correct if proposition $H_{1(2)}$ is true. Conversely, decision $d_{1(2)}$ is not correct if proposition $H_{1(2)}$ is *not* true. A loss function suitable to describe such a two-action decision problem is the '$0 - l_p$' loss function, for $p = \{1, 2\}$, as in Table 2.15, where $l_1 = L(d_1, H_2)$ represents the loss of classifying an item of population 2 (proposition H_2 is true) as a member of population 1 (decision d_1 is taken), and $l_2 = L(d_2, H_1)$ represents the loss of classifying an item of population 1 (proposition H_1 is true) as a member of population 2 (decision d_2 is taken). The loss is zero whenever a correct decision is taken, that is, $L(d_1, H_1) = L(d_2, H_2) = 0$. Conversely, whenever an incorrect decision is taken, one incurs a positive loss. The losses l_1 and l_2 may be equal, whenever the incorrect decisions are considered equally undesirable. Otherwise, the magnitude of each loss will represent the undesirability of each specific occurrence.

Table 2.15 The '$0 - l_p$' loss function, for $p = \{1, 2\}$.

	H_1	H_2
d_1	0	l_1
d_2	l_2	0

Decisions d_1 and d_2 refer to the classification of an observation on an unknown item into population 1 (H_1) and 2 (H_2), respectively

For each decision, one can now compute the expected loss EL(·) (2.19) as follows:

$$\mathrm{EL}(d_1) = \underbrace{L(d_1, H_1)}_{0} \Pr(H_1 \mid E) + \underbrace{L(d_1, H_2)}_{l_1} \Pr(H_2 \mid E) = l_1 \alpha_2, \quad (2.20)$$

$$\mathrm{EL}(d_2) = \underbrace{L(d_2, H_2)}_{0} \Pr(H_2 \mid E) + \underbrace{L(d_2, H_1)}_{l_2} \Pr(H_1 \mid E) = l_2 \alpha_1. \quad (2.21)$$

The best decision after having observed E will be the decision that minimises the expected loss that is calculated using the probabilities on the uncertain events or states of nature as presented in Section 2.8.1.

The plant of unknown origin is classified in population 1 (drug type) if the expected loss of decision d_2 (2.21) is greater than the expected loss of decision d_1 (2.20), that is if

$$\mathrm{EL}(d_2) = l_2 \alpha_1 > l_1 \alpha_2 = \mathrm{EL}(d_1). \quad (2.22)$$

Otherwise, if the expected loss of decision d_1 is greater than the expected loss of decision d_2, the questioned plant is classified in population 2 (fibre type). Rearranging terms in (2.22) as $\alpha_1/\alpha_2 > l_1/l_2$ and dividing both sides by the prior odds in favour of H_1, a threshold k for

the interpretation of the Bayes' factor can be obtained, that is,

$$\text{BF} = \frac{\alpha_1}{\alpha_2} / \frac{\pi_1}{\pi_2} > \frac{l_1}{l_2} / \frac{\pi_1}{\pi_2} = k. \tag{2.23}$$

The optimal decision criterion is to classify the observation in population 1(2) whenever the Bayes' factor is greater (lower) than k. Note that whenever it is reasonable to adopt a symmetric loss function, and when equal *a priori* probabilities for the states of nature are assumed, then the decision criterion is to classify the available observation in population 1(2) whenever the Bayes' factor is greater (less) than 1. An example of such a procedure is also presented in Biedermann et al. (2017).

A review of decision theory in the law and forensic science can be found in Taroni et al. (2020). Applications of statistical decision theory to cases of disputed kinship can be found in Taroni et al. (2005, 2007). Questions related to individualization in forensic science are discussed in Biedermann et al. (2008a, 2016a). Gittelson et al. (2012b, 2013b, 2014, 2016b), addressed the problems of fingermark selection, database selection, genotype designation, and the use of replicates, respectively. Applications of decision theory using probabilistic graphical models, i.e. Bayesian (decision) networks (influence diagrams), can be found in Sections 2.9.3 and 4.7.2 and in Gittelson (2013), Taroni et al. (2014a).

The approach described in this Section can be extended to continuous states of nature θ as described in Section 4.7.

2.9 GRAPHICAL PROBABILISTIC MODELS: BAYESIAN NETWORKS

Methods of formal reasoning have been proposed to assist forensic scientists and lawyers as they seek to improve the understanding of the dependencies which may exist between different aspects of evidence and to deal with the formal analysis of decision making. One very common diagrammatic approach uses graphical probabilistic models and in particular *Bayesian networks* (BNs). These have been found to provide a valuable aid in the representation of relationships amongst characteristics of interest in situations of uncertainty, unpredictability or imprecision.

The use of graphical models to represent and analyse selected aspects of legal cases is not new. Non-probabilistic charting methods developed by Wigmore (1937) can be taken as a predecessor of modern graphical methods such as BNs. Examples of the use of such charts, which were developed to provide formal support for the conclusions reached based on many pieces of evidence, can be found in Robertson and Vignaux (1993b), Schum (1994), Anderson and Twining (1998),

and Roberts and Aitken (2013). For an approach called 'route diagrams', which formally includes probability, see Friedman (1986a) and Friedman (1986b).

Probabilistic networks have attracted the attention of researchers for reviewing the analyses of complex and famous cases such as the Collins case (Edwards, 1991) (see also Section 3.4) and the Sacco and Vanzetti case (Kadane and Schum, 1996), with an emphasis on the credibility and relevance of testimonial evidence. Aspects of the Omar Raddad case (Levitt and Laskey, 2001), the O.J. Simpson trial (Thagart, 2003), and the Busetto case (Taroni et al., 2018b) have also been analysed using graphical models. Relevant analyses and discussions are presented in Schum (1994, 1999). A discussion of the use of Bayesian networks in international criminal trials is presented in McDermott and Aitken (2017).

Among forensic scientists the interest in Bayesian networks stems from the need for a coherent approach to the evaluation of scientific evidence (Evett et al., 2002a; Juchli et al., 2012; Taroni et al., 2014a; Taylor et al., 2016c, 2017b, 2018a,c; Cereda et al., 2018; Sironi et al., 2018).

For illustrative purposes emphasis is placed here on simple networks for the topics of one stain-one offender (Section 2.9.2.3) and the potential of error (e.g. false positives, Section 2.9.2.3). The discussion will focus on determination of the factors (nodes), relevance relationships (arcs), and probabilities to be included.

2.9.1 Elements of the Bayesian Networks

The name 'Bayesian networks' is one of several designations that are commonly encountered, depending on the field of application. Typical variations of the name are 'Bayes nets', 'Bayesian belief networks', 'Bayesian expert systems', 'graphical probabilistic networks', 'probabilistic network models', or 'causal networks', where there may be nuances in the details of definition. For example, the term 'belief' in 'Bayesian belief networks' emphasises that probabilistic assignments in a given model reflect degrees of personal belief. In turn, the notion of an expert system may refer to a broader system of which a graphical probabilistic model is only one element among others. In yet other contexts, the notion of 'causality' is sometimes used. This stems from the fact that the arcs connecting nodes in a Bayesian network can be interpreted as causal relationships, even though the definition of Bayesian networks does not refer to causality and there is no requirement to consider arcs representing assumed causal impact.

More formally, Bayesian networks are graphical probabilistic models that combine probability theory and graph theory. They provide a sophisticated concept for dealing with uncertainty and complexity that represent challenges that occur throughout applied mathematics and engineering (e.g. Jordan, 1999).

Methodologically, the idea of a graphical model is to combine simpler parts in order to ease the approach to a broader inference problem. Probability theory ensures that the component parts are combined suitably and that the system as a whole is coherent, allowing sound inferences to be made. As such, Bayesian networks help to specify relevant probabilistic formulae without displaying their full algebraic form, and make the required probability calculations almost completely automated. The practical implementation of such models can be supported by widely available software solutions. They provide assistance in the intellectually difficult task of organising and arraying complex sets of evidence, and exhibiting their dependencies and independencies in a visual and intuitive way. In essence, the definition of a Bayesian network covers the following elements:

- A finite collection of random variables that are represented by nodes. Each of these nodes has a finite set of mutually exclusive states. In the context, these are also sometimes called 'outcomes'.
- A set of directed edges that connect pairs of nodes.
- The set of variables and the set of directed edges are combined in such a way that a directed acyclic graph is obtained, that is, a graph where no loops are permitted.
- Node probability tables are associated with each variable of a network. The probability table of a variable A that receives entering edges

from variables $B_1, ..., B_n$ contains conditional probabilities $\Pr(A \mid B_1, ..., B_n)$, whereas a variable A with no entering edges from other variables contains unconditional probabilities $\Pr(A)$. Note that the term 'unconditional' refers here only to the absence of an explicit conditioning on other variables (nodes) in a network. Strictly speaking, a probability of the kind $\Pr(A)$ is also considered as conditional because there is always contextual information, habitually denoted by the letter I, which is used when quantifying $\Pr(A)$ (see Section 1.7.5). Thus, $\Pr(A)$ should be written as $\Pr(A \mid I)$, but the letter I is often omitted for ease of notation. Note also that if there is an edge pointing from node A to node B it is said that A is a *parent* of B and B is a *child* of A. A node with no parent is called a root node.

The nodes of a Bayesian network represent propositional or evidential variables of interest. These are, broadly speaking, statements or assertions that such and such is the case. This may be an outcome or a state of nature. It is assumed that personal degrees of belief can be assigned to these states. Propositions are basic intellectual attributes formed by an individual during the course of a reasoning task. A proposition can be thought of as referring to states of affairs. Most often, the actual state may not be known with certainty. For example, there may be uncertainty about the truth or otherwise of the proposition

according to which a crime stain has been left by the offender (i.e. whether or not a stain is relevant). Within a Bayesian network, such a proposition is conceptualised in terms of a node, whose states represent the truth and the falsity of that proposition, respectively. The degree of belief maintained in each of these states is expressed numerically, that is, in terms of probabilities. These probabilities are organised to form the node's probability table.

The totality of the mutually exclusive states of a variable is also referred to as the 'domain' of the variable. The domain of a variable may take one of different forms that will determine the variable's subtype. Examples include {red, green, blue} for a labelled variable, $\{T, F\}$ for a Boolean variable (with 'T' and 'F' denoting 'true' and 'false', respectively), $\{1,2,4,5,7,8\}$ for an integer-numbered variable, and $\{(-1;0],(0;10],(10;100]\}$ for an interval node. Note that the last node type is usable for specifying the intervals over which a continuous quantity can be considered in a discrete form.

The arrows in a Bayesian network represent relevance relationships that an expert modeller assumes to hold within the context of an inferential problem at hand. If a network is properly constructed, then a directed edge from a node A to a node B signifies that A has a direct influence on B. The links between nodes are

sometimes interpreted as 'causal relationships' but the definition of Bayesian networks does not require that the links represent causal impact. Generally, the links in a network are considered to represent probabilistic relevance relationships.

A distinctive feature of Bayesian networks is their incorporation of probability in terms of tables, associated with each node. This allows for interpreting the nature and the strengths of the relationships between a network's different graphical components. Node tables can accommodate probabilities from a variety of different sources. Among the most common sources are personal probabilities from human experts and (statistical) data from databases or literature. Node probability tables can thus be considered as a means of interfacing a model to data. Besides, some Bayesian network programs support the specification of node probability tables through the use of mathematical expressions by exploiting various variable subtypes as described earlier.

The combination of nodes and arrows form paths and create what is known as a net. Therefore, a net can be taken as a compact graphical representation of an evolution of all possible versions of a given case. Early examples of the use of Bayesian networks in forensic science have been given in Aitken and Gammerman (1989), Aitken et al. (1996a,b), Dawid and Evett (1997), Dawid et al. (2002), Evett et al. (2002a), Garbolino and Taroni (2002), Aitken et al. (2003) and Mortera

et al. (2003). These contributions highlighted the following key advantages:

- the ability to structure inferential processes, permitting the consideration of problems in a logical and sequential fashion;
- the requirement to evaluate all possible stories;
- the communication of the processes involved in the inferential problems to others in a succinct manner, illustrating the assumptions made at each node;
- the ability to focus the discussion on probability and underlying assumptions.

2.9.2 The Construction of Bayesian Networks

When constructing Bayesian networks it is important to keep in mind that they do not represent the flow of information, but serve as a direct representation of a part of the real world (Jensen, 2001). This means experts use Bayesian networks to articulate their personal view of a real-world system both graphically and numerically. The result of the modelling process will mainly be influenced by the properties and the experts' individual views, perception, and, ultimately, extent of understanding, of the domain of interest. The problem is well posed in Dawid et al. (2002) where the authors argued that finding an appropriate representation of a case

under examination is crucial for several reasons (viability, computational routines, etc.), and that the graphical construction is to some extent an art-form, but one that can be guided by scientific and logical considerations. The search for good representations for specific problems therefore is an important task for continuing research in this area and was approached by Korb and Nicholson (2011).

Thus, the appropriateness of a given Bayesian network should be assessed with respect to the context in which its construction took place. For example, there may be situations in which the knowledge about a problem domain is severely limited. In addition, there may be processes taking place that are incompletely understood and apparently random (e.g. certain phenomena of transfer of trace materials). Furthermore, the imperfect domain knowledge may be impossible to improve, or may only be improved at an unacceptably high cost. Notwithstanding this, it is possible to view Bayesian network construction as a continuous and adaptive process. A Bayesian network can be taken as an instant representation of a given state of knowledge about a problem of interest. As new knowledge becomes available, the qualitative (net structure) and/or quantitative specifications may be revised in order to account for the newly acquired understanding of domain properties.

Concerning the recurrent question 'Is there a true model?' Lindley (2000) argued that

> *A model is merely your reflection of reality and, like probability, it describes neither you nor the world, but only a relationship between you and that world. It is unsound to refer to the true model. (p. 303)*

It should also be observed that different models can serve to represent questions surrounding the same problem, because (i) the same problem can be approached at different levels of detail and, (ii) existing opinions about domain properties may diverge, as noted, for example, by Garbolino (2001):

> *[...] either you agree with me that E is relevant for H, but our likelihoods are different, or you believe that E is directly relevant for H, and I believe that it is only indirectly relevant, or you believe that it is relevant and I believe it is not. These disagreements explain why we can offer different Bayesian networks models for the same hypothesis. (p. 1506)*

There are three basic types of connections among nodes in a Bayesian network: *serial*, *diverging*, and *converging* connections. These are illustrated in Figure 2.1. There is a serial connection associating three nodes A, B, and C when there is an arrow from A to B, another one from B to C and no arrow from A to C. A serial connection is appropriate when we judge that knowledge of the truth-state of A provides relevant information about the occurrence of B and knowledge of the truth-state of B provides in turn relevant information about C but, when the truth-state of B is known, then knowledge of the

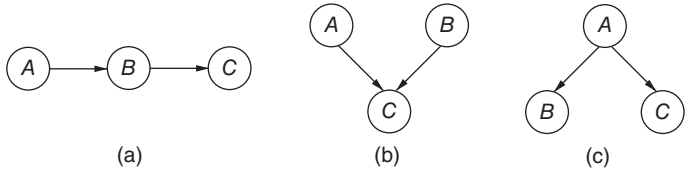

Figure 2.1 Basic connections in Bayesian networks: (a) serial, (b) converging, and (c) diverging connection.

state of A does not provide relevant information about C any more. A influences C through B but only B directly influences C or, in other words, B 'screens off' C from A. If the value of B is known, then A and C are probabilistically independent, i.e. $\Pr(C \mid A, B) = \Pr(C \mid B)$. A serial connection among three nodes is the simplest example of what is known as a *Markov chain*.

As an example, let A be the proposition 'The PoI is the offender', B the proposition 'The bloodstain found on the crime scene comes from the PoI', and C 'The PoI's blood sample and the bloodstain from the crime scene share the same DNA profile'. Then A is relevant for B and B for C but, given B, the cause of the presence of blood could be different from A.

An example of a converging connection with the three nodes A, B, and C applies when there is an arrow from A pointing to C and another one from B pointing to C, with no arrow between A and B. It is said that A and B are probabilistically independent unless either the value of C or the value of any of the children of C is known. Another way of expressing the same idea is to say that A and

B are conditionally dependent given the value of C. Thus $\Pr(AB) = \Pr(A)\Pr(B)$ but $\Pr(AB \mid C)$ may not be equal to $\Pr(A \mid C)\Pr(B \mid C)$. Contrast this with the discussion in Section 1.7.10 where there are events which are conditionally independent but not unconditionally independent.

For example, let A be 'The PoI is the offender' and B 'The bloodstain found on the scene of the crime comes from the offender'. Knowledge that one of these events occurred would not provide information about the occurrence of the other, but if C ('The bloodstain found on the crime scene comes from the PoI') is true, then A and B become related. Converging connections in Bayesian networks are particularly important because they represent a very common pattern of reasoning called conditional dependence or 'explaining away'.

An example of a diverging connection with three nodes A, B, and C holds when there are two arrows originating from A and pointing to B and C, and there is no arrow between B and C. Here A separates B from C: if the value of A is known, then B and C are probabilistically independent, i.e. $\Pr(B \mid A, C) = \Pr(B \mid A)$ and $\Pr(C \mid A, B) = \Pr(C \mid A)$. A diverging connection is the graphical representation of what may be called a spurious correlation. Nodes B and C are correlated because they both depend on a third factor, A. When A is fixed, the correlation vanishes. There are many examples of such spurious correlations. For example, a positive correlation may be shown between the number of doctors in a town and the

number of deaths in a town. As the number of doctors (B) increases, so does the number of deaths (C). This does not mean that doctors are bad for one's health. Rather it is the case that both factors are correlated positively to the population of the town (A). Another example can be considered by interpreting node A as the proposition 'the PoI has assaulted the victim', B as the proposition 'the bloodstain on the clothes of the PoI comes from the victim', and C as the proposition 'the bloodstain on the victim comes from the PoI'.

2.9.2.1 d-Separation Properties

The criterion of *d*-separation, where *d* denotes 'directional', is a graphical criterion (Pearl, 1988) that designates the blocking (or stopping) of the flow of information (or of dependencies) between variables that are connected through a path. Consider this concept through the three basic connections, serial, diverging, and converging (Figure 2.1) that are possible in Bayesian networks:

- In serial and diverging connections, two nodes are said to be *d*-separated if the middle variable is instantiated (i.e. its state is changed from unknown to known),
- In converging connections, two nodes are called *d*-separated as long as the intermediate variable, or one of its descendants, is not instantiated.

Stated otherwise, if two variables in a network are *d*-separated, then changes in the truth state of one variable will have no impact on the truth state of the other variable. If two variables are not *d*-separated, they are called *d*-connected.

2.9.2.2 Chain Rule for the Bayesian Network

Through the use of Bayes' theorem, every joint probability distribution can be decomposed into a product of conditional probabilities. However, the joint probability table grows exponentially with the number of terms. This complexity can be reduced when working with Bayesian networks where it is supposed that a variable, given knowledge of its parents, is independent of all other variables. If the conditional relationships implied by the structure of a Bayesian network hold for a set of variables $A_1, ..., A_n$, then the joint probability distribution $\Pr(A_1, ..., A_n)$ is given by the product of all specified conditional probabilities

$$\Pr(A_1, ..., A_n) = \prod_{i=1}^{n} \Pr(A_i \mid \text{par}(A_i)),$$

where $\text{par}(A_i)$ denotes the set of parental variables of A_i.

Consider the chain rule in case of the three basic connections that are possible in Bayesian networks (Figure 2.1). For a path from A to C

via B, as shown in Figure 2.1(i), $\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A, B)$ can be reduced to $\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid B)$. For a diverging connection, the joint probability can be written as $\Pr(C, A, B) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A)$, whereas in a converging connection it would be $\Pr(A, B, C) = \Pr(A)\Pr(B)\Pr(C \mid A, B)$.

2.9.2.3 Bayesian Network Examples

Consider two simple examples that illustrate how Bayesian networks provide a method for decomposing a joint probability distribution of multiple variables into a set of local distributions of a few variables within each set.

Example 2.8. Consider a proposition E, referring to the correspondence between characteristics of the recovered stain and a DNA profile of a PoI (i.e. evidence), and the source-level proposition H according to which the PoI (or, an unknown person) is the donor of the recovered stain. The relationship between these two propositions is shown in Figure 2.2. Note that H is the name of the root node in this network, whereas the states underlying this node, H_p and H_d (not directly shown in Figure 2.2), refer to whether the stain comes from the PoI or an unknown person, respectively.

The directed arrow from H to E indicates that probabilities are available for $\Pr(E \mid H_p)$ and $\Pr(E \mid H_d)$. It is desired to determine $\Pr(H_p \mid E)$.

Figure 2.2 Simple two-node Bayesian network for a proposition E, relating to a scientist's observation (i.e. evidence) of corresponding features between questioned and known materials, and propositions H referring to common source.

Given values for $\Pr(H_p)$, $\Pr(E \mid H_p)$, and $\Pr(E \mid H_d)$, the posterior probability $\Pr(H_p \mid E)$ is obtained using Bayes' theorem in the usual way:

$$\Pr(H_p \mid E) = \frac{\Pr(E \mid H_p) \times \Pr(H_p)}{\Pr(E \mid H_p) \times \Pr(H_p) + \Pr(E \mid H_d) \times \Pr(H_d)}.$$

At this point, the purpose is solely to illustrate how Bayesian networks support probabilistic computations in principle. Clearly, for simple situations involving only two binary nodes, sophisticated graphical models may not be necessary. However, as soon as multiple nodes, with possibly complex relevance relationships, need to be considered Bayesian networks prove to be helpful. The next example provides an illustration of the gradual increase in complexity arising from the addition of further nodes.

Example 2.9. This example concentrates on evidence defined in terms of a reported match (RM) that is the scientist's assertion of a correspondence observed between the DNA profile of a bloodstain found on the clothing of a victim of a

crime and the DNA profile of a PoI. This definition seeks to provide a closer account of the nature of the scientist's report, which may be erroneous (see Section 6.1.6.4). In other words, when a scientist reports corresponding features, this does not imply that there are in fact corresponding features between the compared items. Thus, a reported match is to be distinguished from the event of a true match (M) between the two compared profiles, an event that itself depends on the proposition H according to which the PoI (H_p), or an unknown person (H_d) is the source of the bloodstain found on the victim's clothing. The relationship between the three nodes RM, M, and H is as shown in Figure 2.3, and is an example of a serial connection.

The arrows from H to M and from M to RM show that probabilities need to be specified for $\Pr(M \mid H_p)$, $\Pr(M \mid H_d)$, $\Pr(\text{RM} \mid M)$, and $\Pr(RM \mid \bar{M})$, where $\bar{M}$ is the complement of M and denotes 'no match'. Note also that the separation of the node RM from the node for H by the node for M shows that RM is conditionally independent of H, given M. In an analogy to Example 2.7, given values for $\Pr(M \mid H_p)$, $\Pr(M \mid H_d)$, $\Pr(\text{RM} \mid M)$,

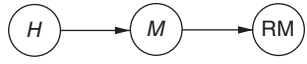

Figure 2.3 Bayes' network for a serial connection for a RM in a DNA profile, where M denotes a match and H a proposition.

$\Pr(\text{RM} \mid \bar{M})$ and $\Pr(H_p)$, it is possible to use Bayes' theorem to determine $\Pr(H_p \mid \text{RM})$.

$$\Pr(H_p \mid \text{RM}) = \frac{\Pr(\text{RM} \mid H_p) \times \Pr(H_p)}{\begin{array}{r}\Pr(\text{RM} \mid H_p) \times \Pr(H_p) \\ + \Pr(\text{RM} \mid H_d) \times \Pr(H_d)\end{array}},$$

where

$$\begin{aligned}\Pr(\text{RM} \mid H_p) &= \Pr(\text{RM} \mid M, H_p) \times \Pr(M \mid H_p) \\ &\quad + \Pr(\text{RM} \mid \bar{M}, H_p) \times \Pr(\bar{M} \mid H_p) \\ &= \Pr(\text{RM} \mid M) \times \Pr(M|H_p) \\ &\quad + \Pr(\text{RM} \mid \bar{M}) \times \Pr(\bar{M}|H_p)\end{aligned}$$

with a similar expression for $\Pr(\text{RM} \mid H_d)$. More details on this practical example are presented in Sections 6.2.6 and 6.3.5. More elaborate diagrams can be analysed in a similar manner known as probability propagation, though the procedures become more complicated. Specific software packages are available that implement efficient algorithms, called updating algorithms, to compute target probabilities.

From the previous examples, the following two intuitive principles can be inferred (Taroni et al., 2004):

- The 'cause' produces the 'effect', i.e. knowing that the 'cause' happened, it can be foreseen that the 'effect' will or might probably occur. This is a predictive line of reasoning, along the direction indicated by a network's arcs.

- The 'effect' does not produce the 'cause', but knowing that the 'effect' occurred, it may be inferred that the 'cause' probably occurred. This is a line of reasoning against the causal direction and one that is diagnostic in nature.

Various applications of Bayesian networks are discussed in Sections 4.4.1, 4.4.2, 4.7.2, and 5.4.

2.9.3 Bayesian Decision Networks (Influence Diagrams)

Bayesian networks can be extended to model decision problems by adding nodes for decisions and utilities. This extended modelling formalism is known as that of Bayesian decision networks. Historically, their development was preceded by decision trees, a method introduced in the 1960s by Raiffa and Schlaifer (1961) to provide a detailed description of decision problems in terms of a graphical mapping of the various decision paths and their outcomes, including probabilities and expected values (i.e. utilities or losses). Bayesian decision networks, also called influence diagrams (Howard and Matheson, 1984; Shachter, 1986), are more compact representations. They take the form of a directed acyclic graph with three types of nodes: chance nodes, decision nodes, and utility nodes. Simple examples are presented in Taroni et al. (2005) and Taroni et al. (2006).

Chance nodes are represented by circles and are the same as the nodes used in Bayesian networks. Chance nodes have a finite set of exhaustive and mutually exclusive states, representing events or states of nature. In turn, decision nodes are represented by square boxes. This type of node has a finite set of feasible alternative decisions (or actions). Utility nodes are the third type of node and are represented by diamond boxes. Utility nodes have no states but to each utility node is associated a utility function over its parents: it represents the expected utility given the states of its parents.

When constructing Bayesian decision networks, it should be noted that utility nodes have no children and that if there is more than one decision node, then there is a directed path consisting of all decision nodes: this ensures that there is a temporal sequence of decisions. Arrows pointing to decision nodes do not carry quantitative information. They only indicate that the state of the decision node's parents is known prior to the decision. Forensic applications of this property are discussed in Gittelson et al. (2013a) and Gittelson (2013).

It should also be observed that decision nodes do not have probability tables associated with them, because it is not meaningful to assign probabilities to a variable under the control of the decision-maker. Chance nodes have, as in Bayesian networks, an associated conditional probability table. However, in Bayesian decision

networks, the arguments of the table can also be decision nodes, not only chance nodes.

The logic of Bayesian decision networks can be illustrated through the philosophical distinction between inference and decision (Lindley, 2000). This distinction asserts that inference and decision are closely related, notably, that the former represents the starting point of the latter: singling out and considering a particular proposition (i.e. state of nature) as true is a decision that depends on one's strength of belief in the truth of the proposition of interest. One's decision also depends on the losses (or utilities) that measure the relative desirability of the consequences given by the combination of choices and individual propositions (that is, actual states of nature).

To clarify these notions, consider the example of forensic individualization (Biedermann et al., 2016a). Consider two discrete, mutually exclusive and exhaustive propositions, such as 'the recovered trace comes from the person of interest (PoI)' (H_p), and 'the trace comes from an unknown person' (H_d). Note that the development can also be made for situations with multiple propositions (Biedermann et al., 2008a, 2016a). In turn, let the available decisions be D_p and D_d, short for 'the PoI is the source of the trace' and 'an unknown person is the source', respectively. A Bayesian decision network for this setting is shown in Figure 2.4. The pair of propositions is represented by the discrete chance node H, in the form of a circle, as introduced earlier in Figure 2.2. It is also supposed

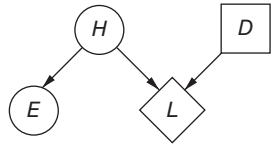

Figure 2.4 Bayesian decision network for analysing the problem of deciding between competing discrete propositions (node H), given evidence E, as an extension of the Bayesian network described in Figure 2.2. The node D represents the available decisions (the number of states equals the number of competing propositions assumed by the node H). The node L accounts for the losses associated for each decision D given each state of the node H.

that there is an item of evidence E that a decision maker judges relevant for judging the truth state of H. The two available decisions are represented by the node D, in the form of a square. The two nodes H and D share a common, diamond-shaped child node L. This node contains the loss function L. Assume, for example, that erroneously deciding D_p, i.e. concluding that the PoI is the source of the crime stain when in fact an unknown person is the source (i.e. H_d is the case) is ten times more serious than erroneously concluding D_d, i.e. concluding that an unknown person is the source when in fact the PoI is the source (i.e. H_p is true). Assume further that correct decisions, that is choosing D_p when H_p holds and D_d when H_d holds, incur a zero loss. This loss function, with l representing the loss associated with a false identification (i.e. choosing D_p when H_d is true), is specified in Table 2.16, associated with the

Table 2.16 Loss function for the Bayesian decision network shown in Figure 2.4.

H:	H_p		H_d	
D:	D_p	D_d	D_p	D_d
L:	0	$l/10$	l	0

node L. In combination, the three nodes D, H, and L represents the decision theoretic core of the network.

Note that the network fragment $H \rightarrow E$ accounts for Bayesian inference. Based on evidence on the node E, a posterior distribution for H (over H_p, H_d) is obtained, given by Bayes' theorem: $\Pr(H \mid E, I) = [\Pr(E \mid H, I) \Pr(H \mid I)] / \Pr(E \mid I)$. The fact that the node H is part of the network that accounts for Bayesian inference ($E \leftarrow H$) as well as part of the network that covers the decision theoretic aspect (i.e. $H \rightarrow L \leftarrow D$), clarifies the idea that the topics of inference and decision are related.

This connection is also highlighted in the expressions of the expected decision losses $\mathrm{EL}(\cdot)$. In particular, the probabilities obtained for the node H enter calculations of the expected losses of the decisions D_p and D_d that, in general, are given by:

$$\begin{aligned}\mathrm{EL}(D_p) = {} & L(D_p, H_p) \Pr(H_p \mid E, I) \\ & + L(D_p, H_d) \Pr(H_d \mid E, I)\end{aligned}$$

and

$$\begin{aligned}\mathrm{EL}(D_d) = {} & L(D_d, H_p)\Pr(H_p \mid E, I) \\ & + L(D_d, H_d)\Pr(H_d \mid E, I).\end{aligned}$$

Thus, for a loss function as defined by Table 2.16 one has

$$\mathrm{EL}(D_p) = L(D_p, H_d)\Pr(H_d \mid E, I) = l\Pr(H_d \mid E, I) \tag{2.24}$$

and

$$\mathrm{EL}(D_d) = L(D_d, H_p)\Pr(H_p \mid E, I) = \frac{l}{10}\Pr(H_p \mid E, I). \tag{2.25}$$

Modern software packages allow such computations for the model shown in Figure 2.4, and other more complex models, to be made in effortless ways.

More generally, note that decision D_p (i.e. identifying the PoI as the source of the crime stain) is the decision with the lower expected loss than D_d (i.e. considering that an unknown person is the source), and hence the optimal decision, when from (2.22):

$$\frac{\Pr(H_p \mid E, I)}{\Pr(H_d \mid E, I)} > \frac{L(D_p, H_d)}{L(D_d, H_p)}. \tag{2.26}$$

This results shows that it is optimal to identify the PoI as the source of the crime stain if and only if the posterior odds are greater than the ratio of the losses associated with two ways in which one's

decision may be erroneous (i.e. the loss of a false identification compared with the loss of a false non-identification).

Further examples of the use of Bayesian decision networks in particular areas of application are presented in Champod et al. (2016b) for identification based on fingermarks and Gittelson et al. (2014) for genotype designation in low-template DNA profiles.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

3

Historical Review

3.1 EARLY HISTORY

The earliest use of probabilistic reasoning in legal decision making, albeit in a somewhat rudimentary form, appears to have been over 18 centuries ago by Jewish scholars in Babylon and Israel writing in the *Talmud* ((Zabell, 1976), in a review of (Rabinovitch, 1973)).[1] For example, if nine stores in a town sold kosher meat and one sold non-kosher meat then a piece of meat which is found, at random, in the town is presumed to be kosher and thus ritually permissible to eat since it is assumed to have come from one of the shops in the majority (Rabinovitch, 1969). However, consider the following quote from the *Talmud*.

> *All that is stationary (fixed) is considered half and half. . . . If nine shops sell ritually slaughtered meat and one sells meat that is not ritually slaughtered and he bought in one of them and does not know which one, it is prohibited because*

[1]For readers interested in the history of probability in general, refer to Franklin (2016).

> *of the doubt; but if meat was found in the street, one goes after the majority. (Kethuboth 15a, quoted in Rabinovitch (1969), p. 438)*

The reasoning seems to be as follows. If the question arises at the source of the meat (i.e. in the shops), the odds in favour of kosher meat are not really 9 to 1. The other nine shops are not considered – the piece of meat certainly did not come from any of them. There are, hence, only two possibilities: the meat is either kosher or it is not. The odds in favour of it being kosher are evens. However, if meat is found out with the shops (e.g. in the street), the probability that it came from any one of the 10 shops is equal for each of the 10 shops. Thus, the probability that it is kosher is 0.9.

The works of Cicero (*De Inventione* and *Rhetorica ad Herennium*) and Quintillian (*Institutio Oratoria*) amongst others are cited by Garber and Zabell (1979). Garber and Zabell also quote an example from Jacob Bernoulli's *Ars Conjectandi* (Bernoulli (1713), Part 4, Chapter 2), which is of interest given the examples in Section 1.7.11 of updating probabilities given new evidence.

The discussion by Bernoulli is as follows. One person, Titius, is found dead on the road. Another, Maevius, is accused of committing the murder. There are various pieces of evidence in support of this accusation.

(1) It is well known that Maevius regarded Titius with hatred. (This is evidence of motive: hatred could have driven Maevius to kill.)

(2) On interrogation, Maevius turned pale and answered apprehensively. (This is evidence of effect: the paleness and apprehension could have come from his own knowledge of having committed a crime.)

(3) A blood stained sword was found in Maevius' house. (This is evidence of a weapon.)

(4) On the day Titius was slain, Maevius travelled over the road. (This is evidence of opportunity.)

(5) A witness, Gaius, alleges that on the day before the murder he had interceded in a dispute between Titius and Maevius.

Later (Chapter 3 of Part 4 of *Ars Conjectandi*) Bernoulli (1713) discussed how to calculate *numerically* the value that should be afforded a piece of evidence or proof.

> *The degree of certainty or the probability which this proof generates can be computed from these cases by the method discussed in the first part (i.e. the ratio of favourable to total cases) just as the fate of the gamblers in games of chance are accustomed to be investigated. (Garber and Zabell, 1979, at p. 44)*

Garber and Zabell (1979) then go on to say

> *What is new in the* Ars Conjectandi *is not its notation of evidence – which is based on the rhetorical treatment of circumstantial evidence – but its attempt to* quantify *such evidence by means of the newly developed calculus of chances. (p. 44)*

Thus it is that over three hundred years ago, consideration was being given to methods of evaluating evidence numerically.

A long discussion of *Ars Conjectandi*, Part 4, is given by Shafer (1978). The distinction is drawn between pure and mixed arguments. A *pure argument* is one that proves a thing in certain cases in such a way as to prove nothing positively in other cases. A *mixed argument*, on the other hand, is one that proves a thing in some cases in such a way that they prove the contrary in the remaining cases. Shafer discusses an example of this from Part 4 of *Ars Conjectandi*.

A man is stabbed with a sword in the midst of a rowdy mob. It is established by the testimony of trustworthy men who were standing at a distance that the crime was committed by a man in a black cloak. It is found that one person, by the name of Gracchus and three others in the crowd were wearing cloaks of that colour. This is an argument that the murder was committed by Gracchus but it is a *mixed* argument. In one case it proves his guilt, in three cases his innocence, according to whether the murder was perpetrated by himself or one of the other three. If one of those three perpetrated the murder then Gracchus is supposed innocent.

However, if at a subsequent hearing, Gracchus went pale, this is a *pure* argument. If the change in his palour arose from a guilty conscience it is indicative of his guilt. If it arose otherwise it does not prove his innocence; it could be that Gracchus went pale for a different reason but that he is still the murderer.

Shafer (1978) draws an analogy between these two kinds of argument and his mathematical theory of evidence (Shafer, 1976) and belief functions (see Section 1.2). In that theory a probability p is assigned to a proposition and a probability q to its negation, or complement, such that $0 \leq p \leq 1$, $0 \leq q \leq 1$, $p + q \leq 1$. It is not necessarily the case that $p + q = 1$, in contradiction to (2.2). There are then three possibilities

- $p > 0$, $q = 0$ implies the presence of evidence in favour of the proposition and the absence of evidence against it;
- $p > 0$, $q > 0$ implies the presence of evidence on both sides, for and against, the proposition;
- $p > 0$, $q > 0$, $p + q = 1$ (additivity) occurs only when there is very strong evidence both for and against the proposition.

Only probabilities that satisfy the additivity rule (2.2) are considered in this book.

Socrates also discussed the use of probabilistic ideas in law.

> *Men care nothing about truth but only about conviction and this is based on probability*

is a quote given by Sheynin (1974). Sheynin also quotes Aristotle (*Rhetorica*, 1376a, p. 19) as saying

> *If you have no witnesses ... you will argue that the judges must decide from what is probable ... If you have*

> *witnesses, and the other man has not, you will argue that probabilities cannot be put on their trial and that we could do without the evidence of witnesses altogether if we need do no more than balance the pleas advanced on either side.*

Sheynin (1974) also mentions that probability in law was discussed by Thomas Aquinas (*Treatise on Law*, Question 105, Article 2, *Great Books*, volume 20, p. 314) who provides a comment on collaborative evidence.

> *In the business affairs of men, there is no such thing as demonstrative and infallible proof and we must contend with a certain conjectural probability .... Consequently, although it is quite possible for two or three witnesses to agree to a falsehood, yet it is neither easy nor probable that they succeed in so doing; therefore their testimony is taken as being true. (p. 108)*

Jacob Bernoulli (1713) gave a probabilistic analysis of the cumulative force of circumstantial evidence. His nephew, Nicholas Bernoulli (1709), applied the calculus of probabilities to problems including the presumption of death, the value of annuities, marine insurance, the veracity of testimony, and the probability of innocence (see, (Fienberg, 1989)).

Further comment on the combination of evidence and the distinction between the roles of fact-finder and witness was given by Locard (1940) who proposed some inspired guidelines for the interpretation of scientific evidence. These guidelines remain pertinent to scientists and lawyers even today.

> *The physical certainty provided by scientific evidence rests upon evidential values of different orders. These are measurable and can be expressed numerically. Hence the expert knows and argues that he knows the truth, but only within the limits of the risks of error inherent to the technique. This numbering of adverse probabilities should be explicitly indicated by the expert. The expert is not the judge: he should not be influenced by facts of a moral sort. His duty is to ignore the trial. It is the judge's duty to evaluate whether or not a single negative evidence, against a sextillion of probabilities, can prevent him from acting. And finally, it is the duty of the judge to decide if the evidence is in that case, proof of guilt. (pp. 286–287)*

The application of probability to the verdicts by juries in civil and criminal trials was discussed by Poisson (1837) and there is also associated work by Condorcet (1785), Cournot (1838), and Laplace (1886). The models developed by Poisson have been put in a modern setting by Gelfand and Solomon (1973).

Two early examples of the use of statistics were to query the authenticity of signatures on wills (Mode, 1963). One of these, the Howland will case from the 1860s has been discussed also by Meier and Zabell (1980). This case is probably the earliest instance in American law of the use of probabilistic and statistical evidence. The evidence was given by Professor Benjamin Peirce, Professor of Mathematics at Harvard University, and by his son Charles, then a member of staff of the United States Coast Survey. The evidence is related to the agreement of 30 downstrokes in a contested signature with those of a genuine

signature. It was argued that the probability of this agreement if the contested signature were genuine was extremely small; the probability of observing two spontaneous signatures with the number of overlaid strokes observed in those two signatures was $(1/5)^{30}$. Hence the contested signature was a forgery. Comments on this case pointed out the now famous *prosecutor's fallacy* (Section 2.5.1). It has also been argued that Charles Peirce (1878) – in a pre-Bayesian statistical model – considered only two hypotheses with implicitly initial odds of 1, thereby excluding some alternatives that might have had a prior probability greater than zero (Good, 1983). Interestingly, in the same article, Good (1983) comments

> *It might be better to call '* ... *hypothesis testing'* hypothesis determination*, as in a court of law where a judge or jury 'determines' that an accused person is innocent or guilty and where stating a numerical probability might even be regarded as contempt of court.' (p. 71)*

More recent attempts to evaluate evidence are reviewed here in greater detail.

3.2 THE DREYFUS CASE

This example concerns the trial of Dreyfus in France at the end of the nineteenth century. Dreyfus, an officer in the French Army assigned to the War Ministry, was accused in 1894 of selling military secrets to the German military attaché.

Part of the evidence against Dreyfus centred on a document called the *bordereau*, admitted to have been written by him, and said by his enemies to contain cipher messages. This assertion was made because of the examination of the position of words in the bordereau.

Quantification with the use of probabilities in this case and others was suggested by Alphonse Bertillon, a Paris police officer who founded a police laboratory for the identification of criminals (Bertillon, 1893, 1898). See, for the sake of illustration, a quote from Bertillon (1898) on the need for a quantification:

> *This writing, characterized by the set of unique features we have enumerated, can only be encountered in one individual among a hundred, among a thousand, among ten thousand or among a million individuals. (p. 20)*

In fact, after reconstructing the bordereau and tracing on it with four millimetre interval vertical lines, Bertillon showed that four pairs of polysyllabic words (among 26 pairs) had the same relative position with respect to the grid. Then, with reference to probability theory, Bertillon stated that the coincidences described could not be attributed to normal handwriting. Therefore, the bordereau was a forged document. Bertillon submitted probability calculations to support his conclusion. His statistical argument can be expressed as follows: if the probability for one coincidence equals 0.2, then the probability of observing N coincidences is $(0.2)^N$. Bertillon

calculated that the four coincidences observed by him had, then, a probability of 0.2^4, or $1/625$, a value that was so small as to demonstrate that the bordereau was a forgery (Charpentier, 1993).

However, this value of 0.2 was chosen purely for illustration and had no evidential foundation; for a comment on this point, see Darboux et al. (1908).

Bertillon's deposition included not only this simple calculation but also an extensive argument to identify Dreyfus as the author of the bordereau on the basis of other measurements and a complex construction of hypotheses. (For an extensive description of the case, see literature quoted in Taroni et al. (1998, p. 189)).

As noted in Section 2.5.1 and using a Bayesian perspective, it is not difficult to see where Bertillon's logic had failed in his conclusion on the forgery. It seems that Bertillon argued that $\Pr(H_d \mid E, I) = p = 1/625 = 0.0016$ and hence that $\Pr(H_p \mid E, I) = 1 - p = 0.9984$. However, p represents $\Pr(E \mid H_d, I)$. This seems to be an early example of the prosecutor's fallacy.

The reliability of Bertillon's approaches were discussed at a retrial. Notably, Darboux, Appell, and Poincaré, mathematicians and members of the French Academy of Sciences, offered their opinions. They commented that the probabilistic assessment proposed by Bertillon had no sound mathematical basis. In fact, the value of 0.0016 is the probability of observing 4 independent coincidences out of 4 comparisons (with the probability, θ, of one coincidence being 0.2),

whereas Darboux, Appell, and Poincaré are quoted as determining the probability of observing 4 coincidences out of 26 comparisons to be quite different, namely, 0.7, or 400 times greater ($0.7/0.0016 = 437.5$) (Moras, 1906; Darboux et al., 1908).

It is not clear how this figure of 0.7 was derived. The binomial expression $\binom{26}{4} 0.2^4 0.8^{22} = 0.176$, and the probability of four or more coincidences out of 26 comparisons is approximately 0.8. It is not possible to choose a value of θ for which $\binom{26}{4} \theta^4 (1-\theta)^{22} = 0.7$. The value of θ for which the probability of four or more coincidences, out of 26, is 0.7 is $\theta = 0.18$. Further comments on Bertillon's calculations are given in Section 3.3.

Another assertion by Dreyfus' enemies was that the letters of the alphabet did not occur in the documents in the proportions in which they were known to occur in average French prose. The proportions observed had a very small probability of occurring; see Tribe (1971). Though it was pointed out to the lawyers that the most probable proportion of letters was itself highly improbable, this point was not properly understood. A simple example from coin tossing will suffice to explain what is meant by the phrase 'the most probable proportion of letters was itself highly improbable'. Consider a fair coin, that is, one in which the probabilities of a head and of a tail are equal at 1/2. If the coin is tossed 10 000 times, the expected

number of heads is 5 000 (see Section A.2.3 with $n = 10\,000, \theta = 1/2$) and this is also the most probable outcome. However, the probability of 5 000 heads, as distinct from 4 999 or 5 001 or any other number, is $\simeq 0.008$ or 1 in 125, which is a very low probability. The most probable outcome is, itself, improbable. The situation, of course, would be considerably enhanced given all the possible choices of combinations of letters in French prose in Dreyfus' time. This idea may be expressed in mathematical symbols as follows. If Dreyfus were innocent (H_d) the positions of the words (E) that he had used would be extremely unlikely; $\Pr(E \mid H_d)$ would be very small. The prosecuting lawyers concluded that Dreyfus must have deliberately chosen the letters he did as a cipher and so must be a spy; $\Pr(H_d \mid E)$ must be very small. The lawyers did not see that any other combination of letters would also be extremely unlikely and that the particular combination used by Dreyfus was of no great significance. This is another example of the prosecutor's fallacy.

Darboux, Appell, and Poincaré also expressed a more fundamental point: the nature of the inferential process they used to reach the conclusion. They stated that the case under consideration was a classical problem of *probability of the causes* and not a problem of *probability of the effects* (Darboux et al., 1908). See Dawid et al. (2016) for recent

debate of the distinction between causes of effects and effects of causes.

The difference between the two statistical concepts (and inferences) could be illustrated by the following example; see again Darboux et al. (1908):

> *If you draw a ball from an urn containing 90 white balls and 10 black balls, the probability of drawing a black ball is 1/10 and it corresponds to the probability of the effect. Suppose now you are facing two identical urns. Urn 1 is containing black and white balls in the proportion 90:10. The second urn contains black and white balls in the proportion 10:90. You choose an urn (at random, each urn is equally likely to be chosen) and pick up a ball. It is white. What is the probability that you have picked up a ball from urn 1? In this example, the effect is known, but it is the cause which is uncertain.*[2] *(p. 502)*

[2]Nous en avons dit assez pour faire comprendre la nécessité d'une base de raisonnement plus solide. C'est ce que les fondateurs du calcul des probabilités ont cherché pour les questions de ce genre, mais nous ne pouvons l'expliquer sans entrer dans quelques détails techniques. Ils ont distingué la probabilité des effets et la probabilité des causes. Comme exemple de probabilité des effets, on choisit d'ordinaire une urne contenant 90 boules blanches et 10 boules noires. Si l'on tire au hasard une boule de cette urne, quelle est la probabilité pour que cette boule soit noire? C'est évidemment 1/10. Les problèmes de probabilité des causes sont beaucoup plus compliqués, mais beaucoup plus intéressants. Supposons par exemple deux urnes d'aspect extérieur identique; nous savons que l'une contient 90 boules blanches et 10 boules noires, et l'autre au contraire 90 boules noires et 10 boules blanches. Nous tirons au hasard une boule de l'une des urnes, sans savoir de laquelle, et nous constatons qu'elle est blanche. Quelle est la

This is another example concerning urns (Section 1.7.2). In order to infer something about a possible cause from an observation of an effect, two assessments are needed: the probabilities *a priori* of the causes under examination (i.e. forgery or not in the Dreyfus case), and the probabilities of the observed effect for each possible cause (the coincidences observed by Bertillon). A more detailed description of this kind of reasoning applied in forensic science was proposed by Poincaré and his colleagues; it will be presented in Section 3.8. See also the comment by Poincaré (1992):

> *An effect may be the product of either cause A or cause B. The effect has already been observed; one wants to know the probability that it is the result of cause A; this is the* a posteriori *probability. But, I'm not able to calculate this if an accepted convention does not permit me to calculate in advance the* a priori *probability for the cause producing the effect; I want to speak of the probability of this eventuality, for one who has never before observed the result. (p. 229)*

For further analyses of the questioned document examination in the Dreyfus' case, see Champod et al. (1999) and Kaye (2007).

probabilité pour que ce soit dans la première urne que nous ayons puisé? Dans ce nouveau problème, l'effet est connu, on a constaté que la boule tirée était blanche; mais la cause est inconnue, on ne sait pas dans quelle urne on a fait le tirage. Le probléme qui nous occupe ici est de même nature: l'effet est connu, ce sont les coincidences signalées sur le bordereau, et c'est la cause (forgerie ou écriture naturelle) qu'il s'agit de déterminer. (Original report at p. 337, reprinted in Darboux et al. (1908) at p. 502)

3.3 STATISTICAL ARGUMENTS BY EARLY TWENTIETH-CENTURY FORENSIC SCIENTISTS

A review of the forensic science literature suggests that, for the evaluation of forensic evidence, early forensic scientists recognised that adequate data and consideration of the case as a whole should be used to reach a decision. Despite the problem encountered by Bertillon during the Dreyfus case (e.g. an early example of what is now known as 'the prosecutor's fallacy', Section 2.5.1), Bertillon can be thought of as the first Bayesian forensic practitioner. In fact, after expressing the need of a quantification of the observed features (Bertillon, 1898), he completed his reasoning by affirming that the only way to accept an expert's categorical conclusions was to consider not only the statistical evidence provided by the examination of the document, but also other information pertaining to the inquiry. He described how the number of people who could be the author of the questioned document size is reduced with the inquiry (i.e. the testimonies and circumstances of the case). This description introduces the general idea of a relevant population, a concept expanded and discussed by Lempert (1991), Robertson and Vignaux (1993b), Champod et al. (2004), Kaye (2004, 2008b), Hicks et al. (2015), Morrison et al. (2016). An important contribution to the role of scientific evidence is that of Fienberg et al. (1996).

He and his co-authors note that (i) what is treated as a relevant population may only be a conveniently available population and (ii) the event that evidence associated with the crime came from the defendant is not necessarily the same as the event that the defendant committed the crime.

Therefore, the evidentiary value of the scientific observations, even if not totally confirmatory of guilt, could supply sufficient information to allow a conviction when the case is considered as a whole. Other examples of similar reasoning were published by Balthazard (1911) and Souder (1934) in the fields of fingermarks and typewritten documents, respectively. A complete historical summary of the relationship of forensic scientists to Bayesian ideas and a demonstration that their points of view were generally compatible with a Bayesian framework is demonstrated in Taroni et al. (1998).

Despite his argument in the Dreyfus case, Bertillon wrote that experts must be prepared to present evidence in a numerical form which was more demanding than that generally required of expert opinions. He proposed that reports should be concluded in the following form (Bertillon, 1898):

> *This writing characterized by the set of unique features the expert enumerated, can only be encountered in one individual among a hundred, among a thousand, among ten thousand or among a million individuals. (p. 20)*

Moreover, Bertillon argued that the only way to accept a conclusion of the final issue (e.g. the identification of a writer) was to consider not only the statistical evidence provided by the examination of the document, but also other information pertaining to the inquiry. Bertillon considered the presentation of results without such information as a methodological error. The value of the comparison results, even if not absolute, could supply sufficient information to allow a conviction when the case is considered as a whole. The same approach – more clearly expressed in a numerical way – was proposed in 1934, for typewritten documents, by William Souder (1934):

> *Suppose the report does not establish an extremely remote possibility of recurrence (of characteristics or agreements between the questioned and known writings). [. . .]. Suppose the final fraction for recurrence of the typed characteristics had come out as only 1 in 100. Is such a report of value? Yes, if the number of typewriters upon which the document could have been written can be limited to 100 or less, the report is vital. Similarly, in handwriting we do not have to push the tests until we get a fraction represented by unity divided by the population of the world. Obviously the denominator can always be reduced to those who can write and further to those having the capacity to produce the work in question. In a special case, it may be possible to prove that one of three individuals must have produced the document. Our report, even though it shows a mathematical probability of only 1 in 100, would then irresistibly establish the conclusion. (pp. 683–684)*

The same idea that a final issue could only be assessed if the case is considered as a whole is often reiterated in judicial literature (Koehler, 1997a).

> *Forensic science alone cannot identify the probability that O.J. Simpson - or any other criminal defendant - is or is not the source of the recovered genetic evidence. Non-genetic considerations must be factored into any equation that purports to identify the chance that someone is the source of genetic sample. (p. 219)*

The debate on the use of statistical thinking in fingerprints/fingermarks has a long history, beginning in 1900. The aim was to assess the probabilistic value for a minute configuration on a given surface size of a fingerprint. As described by Champod (1996), the statistical model evolution begun with E. R. Henry in 1900 with arguments presented in his book on classification and uses of fingerprints (Henry, 1913). A critical analysis of the principal models proposed over a long period of time for the quantitative assessment of fingerprints is presented in Stoney and Thornton (1986). A historical and critical description of such an evolution is also presented in Champod (1995) starting from controversies around the Edmond Locard's 'tripartite rule' and notably on the qualitative aspect presented in Locard (1914). The criterion is the following:

> *If a limited number of characteristic points are present, the fingerprint cannot provide certainty for an identification,*

> *but only a presumption proportional to the number of points available and their clarity.*[3] *(p. 332)*

This criterion implicitly supported the development of quantitative models.

Another old example of the use of a probabilistic inference for fingerprint identification is the work of Balthazard, a French legal examiner. Balthazard's work influenced rules regarding standards for the establishment of an identification in a fingerprint such as the rule for seventeen concordant minutiae expressed in Italian jurisprudence since 1954 (Balthazard, 1911).

Despite the weakness of Balthazard's hypotheses and assumptions used to perform his simple calculation that have been extensively challenged in the scientific literature (see comments in (Champod et al., 2016b)), it is important to note that part of Balthazard's text is in agreement with the Bayesian framework (Balthazard, 1911).

> *In medico-legal work, the number of corresponding minutiae can be lowered to eleven or twelve if you can be certain that the population of potential criminals is not the entire world population but it is restricted to an inhabitant of Europe, a French citizen, or an inhabitant of a city, or of a village, etc. (p. 1864)*

So, as years later Souder proposed for questioned documents, here Balthazard stated that prior assessment (based on inquiry information and a reduction of the size of the suspect population)

[3]Translation reported in Champod (1995), p. 136.

has to be associated with a statistical value of the evidence to allow the decision-maker to judge on an identification (a posterior assessment).

The model's evolution continues to the explicit interest on the Bayesian model and on a likelihood ratio calculation as proposed by Stoney (1985). Note that Lempert (1977) promoted the use of the likelihood ratio for assessing the value of fingermarks in his seminal paper. The same direction has been followed and developed by Champod (1996). More discussion around the use of the Bayesian approach in fingermarks interpretation can be read in Stoney and Thornton (1988) and Kingston (1988) where the interest on the use of other (non-fingerprint) evidence to define a limited suspect population and then using this population to define a prior probability is discussed. Interestingly, Kingston (1988) concluded his comment suggesting the need to generalise the use of such an approach by saying :

> *The model under discussion here [the Bayesian model] has applications to a range of evidence types where individualization is of concern. It is clearly necessary to examine more closely the interpretation. and the presentation, of opinion with respect to physical evidence types which lend themselves to conclusions that are far less certain than is the case for fingerprints. (p. 11)*

Unfortunately, over one hundred years later, the discussion on fingerprint 'identification' and the use of probabilistic models is still open. See, for example, the discussions proposed by Taroni

and Margot (2000), Champod and Evett (2001), Friedman et al. (2002), the recommendations of the US National Institute of Science and Technology report on Expert Working Group on Human Factors in Latent Print Analysis (2012), Champod et al. (2016b) and references therein, Neumann and Stern (2016) and by Swofford et al. (2018).

3.4 PEOPLE *v.* COLLINS

The Dreyfus case is a rather straightforward abuse of probabilistic ideas though the fallacy of which it is an example still occurs. It is easy now to expose the fallacy through consideration of the odds form of Bayes' theorem (see Section 2.3). At the time, however, the difficulty of having the correct reasoning accepted had serious and unfortunate consequences for Dreyfus. A further example of the fallacy occurred in a case that has achieved a certain notoriety in the probabilistic legal literature, namely, that of *People v. Collins* (1968), (Kingston, 1965a,b; Fairley and Mosteller, 1974, 1977; Koehler, 1997a). In this case, probability values, for which there was no objective justification, were quoted in court.

Briefly, the crime was as follows. An old lady, Juanita Brooks, was pushed to the ground in an alley-way in the San Pedro area of Los Angeles by someone whom she neither saw nor heard. According to Mrs. Brooks, a blond-haired woman

wearing dark clothing grabbed her purse and ran away. John Bass, who lived at the end of the alley, heard the commotion and saw a blond-haired woman wearing dark clothing run from the scene. He also noticed that the woman had a ponytail and that she entered a yellow car driven by a black man who had a beard and a moustache.

A couple answering this description were eventually arrested and brought to trial. The prosecutor called as a witness an instructor of mathematics at a state college in an attempt to bolster the identifications. This witness testified to the product rule for multiplying together the probabilities of independent events (the third law of probability (1.10)). If $E_1, E_2, \ldots, E_n$ are mutually independent pieces of evidence and H_d denotes the hypothesis of innocence then the extension of the third law to a set of n independent events is

$$\Pr(E_1, E_2, \ldots, E_n \mid H_d) = \Pr(E_1 \mid H_d) \Pr(E_2 \mid H_d) \cdots \Pr(E_n \mid H_d). \qquad (3.1)$$

In words, this states that, for mutually independent events, the probability that they all happen is the product of the probabilities of each individual event happening.

The instructor of mathematics then applied this rule to the characteristics as testified to by the other witnesses. Values to be used for the probabilities of the individual characteristics were suggested by the prosecutor without any

Table 3.1 Probabilities for various characteristics of the couple observed in the case of *People v. Collins*.

Evidence	Characteristic	Probability
E_1	Partly yellow automobile	1/10
E_2	Man with moustache	1/4
E_3	Girl with ponytail	1/10
E_4	Girl with blonde hair	1/3
E_5	Negro man with beard	1/10
E_6	Interracial couple in car	1/1 000

justification, a procedure that would not now go unchallenged. The jurors were invited to choose their own values but, naturally, there is no record of whether they did or not. The individual probabilities suggested by the prosecutor are given in Table 3.1. Using the product rule for independent characteristics, the prosecutor calculated the probability that a couple selected at random from a population would exhibit all these characteristics as 1 in 12 million ($10 \times 4 \times 10 \times 3 \times 10 \times 1\,000 = 12\,000\,000$).

The accused were found guilty. This verdict was overturned on appeal for two statistical reasons.

(a) The statistical testimony lacked an adequate foundation both in evidence and in statistical theory.

(b) The testimony and the manner in which the prosecution used it distracted the jury from its proper function of weighing the evidence on the issue of guilt.

The first reason refers to the lack of justification offered for the choice of probability values and the assumption that the various characteristics were independent. As an example of this latter point, an assumption of independence implies that the propensity of a man to have a moustache does not affect his propensity to have a beard. Moreover, the computation implicitly assumed that the six reported characteristics were true and were accurately reported. It made no allowance for the possibility of disguise (e.g. dyed hair).

The second reason still has considerable force today. When statistical evidence is presented, great care has to be taken that the jury is not distracted from its proper function of weighing the evidence on the issue of guilt. Care has also to be taken to avoid cognitive bias. This is a procedure by which exposure of the expert to irrelevant information can potentially cause bias in their report (Dror, 2018).

The fallacy of the transposed conditional is also evident. The evidence is $(E_1, E_2, \ldots, E_6)$ and $\Pr(E_1, E_2, \ldots, E_6 \mid H_d)$ is extremely small (1 in 12 million). The temptation for a juror to interpret this figure as a probability of innocence is very great.

Lest it be thought that matters have improved, consider the case of *R v. Clark* (1999, 2000, and 2003). Sally Clark's first child Christopher died unexpectedly at the age of about three months when Clark was the only other person in the house. The death was initially treated as a case

of Sudden Infant Death Syndrome (SIDS). Her second child, Harry, was born the following year. He died in similar circumstances. Sally Clark was arrested and charged with murdering both her children. At trial, a professor of paediatrics quoted from a report (Fleming et al., 2000) that, in a family like the Clarks, the probability that two babies would both die of SIDS was around 1 in 73 million. This was based on a study which estimated the probability of a single SIDS death in such a family as 1 in 8 500, and then squared this to obtain the probability of two deaths, a mathematical operation that assumed the two deaths were independent. In an open letter (dated 23 January 2002) to the Lord Chancellor, copied to the President of the Law Society of England and Wales, the Royal Statistical Society expressed its concern at the statistical errors which can occur in the courts, with particular reference to Clark. To quote from the letter:

> *One focus of the public attention was the statistical evidence given by a medical expert witness who drew on a published study (Confidential Enquiry into Stillbirths and Deaths in Infancy, Fleming et al. (2000)) to obtain an estimate (1 in 8 543) of the frequency of SIDS (or 'cot death') in families having some of the characteristics of the defendant's family. The witness went on to square this estimate to obtain a value of 1 in 73 million for the frequency of two cases of SIDS in such a family.*
>
> *[…] Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis)-interpretation is a serious error of logic known as the prosecutor's fallacy. (Royal Statistical*

Society News, March 2002, reproduced at http://www .rss.org.uk/statsandlaw.)

An interesting discussion on the role of the propositions is given in Dawid (2002) and Hill (2004). Two propositions may be considered:

H_p: the mother did indeed murder two of her babies;

H_d: the babies died of SIDS.

The Bayesian approach considers the relative likelihoods of the evidence E of the two deaths under the two propositions. This is the ratio of *Pr*(two dead babies | murdered) / *Pr*(two dead babies | SIDS deaths). Both probabilities in this ratio are 1: the probability of two dead babies if they have been murdered is 1 and the probability of two dead babies if they have died of SIDS is also 1. For these two propositions, the value of the evidence of the deaths is not relevant. However, the determination of the posterior odds $\Pr(H_p \mid E)/ \Pr(H_d \mid E)$ requires consideration of the prior probabilities for H_p and H_d as well as the likelihood ratio. First, consider H_d: the probability of the two deaths by SIDS is taken to be 1 in 73 million (assuming the figure given in the trial to be accurate). For the probability of death by murder (H_p), Dawid (2002) cites Office of National Statistics (ONS) data for 1997 in which there were 642 093 live births and seven babies were murdered in the first year of life. The probability of being murdered is then estimated by Dawid (2002)

to be $7/642\,093 = 1.1 \times 10^{-5}$ or approximately 1 in 90 000. Assuming independence, as for the analysis assuming SIDS deaths (not a particularly reasonable assumption here either), and squaring the probability for two murders, a probability is obtained of 1.2×10^{-10} or about 1 in 8.4 billion for the probability of two murders of babies in the same family. The ratio $\Pr(H_p)/\Pr(H_d)$ of 1 in 8.4 billion to 1 in 73 million is 1/115. This figure is the prior odds for H_p relative to H_d. As the likelihood ratio has a value of 1, the posterior odds for H_p relative to H_d are also 1 in 115. Thus, the posterior probability that the babies died of SIDS rather than being murdered is 115/116 or 0.99.

It is important that the propositions are specified carefully. Dawid (2002) provided an alternative analysis in which the propositions considered are not so restricted as those above (see Section 5.2 and drew a distinction between *proposition* and *explanation*). In the alternative analysis, let the propositions be

H_p: The babies were murdered;

H_d: The babies were not murdered.

The evidence E is, as in the previous analysis, that the babies died. Here it is further assumed that if the babies were murdered then it was Sally Clark that did the murders and if the babies were not murdered then they died of SIDS. Thus $\Pr(E \mid H_p) = 1$, as before. However, H_d does not include the implication that the babies died. Thus, $\Pr(E \mid H_d)$ is the probability that the babies died,

assuming they were not murdered. This is taken to be the probability that they died of natural causes, and more specifically SIDS. This is then 1 in 73 million (still using the figure provided in the original trial). The likelihood ratio is then 1 divided by 1 in 73 million, or 73 million, a figure that provides very strong evidence in support of the proposition that the babies were murdered by Sally Clark. However, the change in propositions from the initial analysis means that the prior odds also change. In this (second) analysis, $\Pr(H_p) = 1/8.4$ billion using the ONS figures provided earlier. Thus $\Pr(H_p)/\Pr(H_d) \simeq 1 / 8.4$ billion. The combination of the likelihood ratio and the prior odds gives the same posterior odds as before: 73 million divided by 8.4 billion or 1 in 115. These give a posterior probability that the babies were not murdered of 115/116 as before.

The probability of other evidence has also to be assessed under each of the two propositions. The choice of probabilities for the other evidence is subjective. However, the Bayesian approach makes it very clear what features of the evidence should be taken into account and what the effect of the choices of probabilities is.

Note that the Court of Appeal heard new medical evidence in January 2003 and Sally Clark's conviction was overturned. She died in 2007. A commentary on this case is also given in Aitken (2003).

A similar American case where the prosecutor's fallacy is committed in a case concerning SIDS is that of *Wilson v. Maryland* (2002).

Note that the prosecutor's fallacy equates a small probability of finding the evidence on an innocent person with the probability that the person is innocent. More details on such a pitfall of intuition have been presented in Sections 2.5.1 and 2.7.1.

3.5 DISCRIMINATING POWER

3.5.1 Derivation

How good is a method at distinguishing between two samples of material from different sources? If a method fails to distinguish between two samples, how strong is this as evidence that the samples come from the same source? Questions like these and the answers to them were of considerable interest to forensic scientists in the late 1960s and in the 1970s; see, for example, theoretical work in Parker (1966, 1967), Jones (1972) and Smalldon and Moffat (1973). Experimental attempts to answer these questions are described in Tippett et al. (1968) for fragments of paint, in Gaudette and Keeping (1974) for human head hairs, in Groom and Lawton (1987) for shoeprints, in Massonnet and Stoecklein (1999) for paint,

and in Adams (2003) for dental impressions. Other examples are given, for example, by Buzzini and Massonnet (2013, 2015) (textile fibres), Muehlethaler et al. (2014) and Falardeau et al. (2019) (paints), Morris et al. (2017) (cartridges), Krol et al. (2014) (printer inks), and Weyermann et al. (2012) (pen inks).

Two individuals are selected at random from some population. The probability that they are found to match with respect to some characteristic (e.g. blood profile, paint fragments on clothing, head hairs) is known as the *probability of non-discrimination* or the *probability of a match*, PM. The complementary probability, the probability they are found not to match with respect to this characteristic, is known as the *probability of discrimination* (Jones, 1972) or *discriminating power*, DP (Smalldon and Moffat, 1973). The idea was first applied to problems concerning ecological diversity (Simpson, 1949) and later to blood group genetics (Fisher, 1951). See also Jeffreys et al. (1987) for an application to DNA profiles.

Consider a population and a locus in which there are k genotypes, labelled $1, \ldots, k$, and in which the j-th genotype has population proportion p_j, such that $p_1 + \cdots + p_k = 1$ from the second law of probability for mutually exclusive and exhaustive events generalised to k events, see Section 1.5. Two people are selected at random from this population such that their genotypes may be assumed independent. What is the probability of a

match of genotypes between the two people at this locus?

Let the two people be called C and D. Let C_1 and D_1 be the events that C and D, respectively, are of genotype labelled 1. Then

$$\Pr(C_1) = \Pr(D_1) = p_1,$$

and the joint probability of C_1 and of D_1 is given by

$$\Pr(C_1 D_1) = \Pr(C_1) \times \Pr(D_1) = p_1^2$$

by the third law of probability (1.10) applied to independent events. Thus, the probability they are both of genotype 1 is p_1^2. In general, let C_j, D_j be the events that C, D are of genotype j ($j = 1, \ldots, k$). The probability that the individuals selected at random match on genotype j, is given by

$$\Pr(C_j D_j) = \Pr(C_j) \times \Pr(D_j) = p_j^2.$$

The probability of a match on *any* genotype is the disjunction of k mutually exclusive events, the matches on genotypes $1, \ldots, k$, respectively. Let Q be the probability PM of a match. Then

$$\begin{aligned} Q &= \Pr(C_1 D_1 \text{ or } C_2 D_2 \text{ or } \ldots \text{ or } C_k D_k) \\ &= \Pr(C_1 D_1) + \Pr(C_2 D_2) + \cdots + \Pr(C_k D_k) \\ &= p_1^2 + p_2^2 + \cdots + p_k^2, \end{aligned} \tag{3.2}$$

by the second and third laws of probability (1.5) and (1.10); see also (1.1). The discriminating power, or probability of discrimination, is $1 - Q$.

Example 3.1. Consider the frequencies from Table 1.1 where $k = 3$. Then

$$Q = 0.188^2 + 0.321^2 + 0.491^2 = 0.379.$$

The discriminating power $\mathrm{DP} = 1 - Q = 0.621$.

3.5.2 Evaluation of Evidence by Discriminating Power

The approach using discriminating power has implications for the assessment of the value of forensic evidence. If two samples of material (e.g. two blood stains, two sets of paint fragments, two groups of human head hairs) are found to be indistinguishable, it is of interest to know if this is forensically significant. If a system has a high value of Q, it implies that a match between samples of materials from two different sources under this system is likely. For example, if there were only one category, no discrimination would be possible. In such a case, $k = 1$, $p_1 = 1$ and $Q = p_1^2 = 1$. It is intuitively reasonable that a match under such a system will not be very significant. Conversely, if a system has a very low value of Q a match will be forensically significant. Blood grouping data from the Strathclyde region of Scotland for which there is a discriminating power of 0.602 are given by Gettinby (1984) and are used here for sake of illustration. He interpreted this, in the context of blood grouping, to mean that

> *[...] in 100 cases where two blood samples come from different people then, on average, 60 will be identifiable as such. (p. 224)*

Limits for Q may be determined (Jones, 1972). First, note that $p_1 + \cdots + p_k = 1$ and that $0 \leq p_j \leq 1$ $(j = 1, \ldots, k)$ from the first law of probability (1.6). Thus $p_j^2 \leq p_j$ $(j = 1, \ldots, k)$ and so $Q = p_1^2 + \cdots + p_k^2 \leq 1$; i.e. Q can never be greater than 1 (and equal to 1 if and only if one of the p_j's is 1, and hence the rest are 0, as illustrated above). A value of Q equal to 1 implies that all members of the population fall into the same category; the discriminating power is zero.

Now, consider the lower bound. It is certainly no less than zero. Suppose the characteristic of interest (h_0, say) of interest divides the system into k classes of equal probability $1/k$ so that $p_j = p_{0j} = 1/k$ $(j = 1, \ldots, k)$. Then

$$\begin{aligned} Q = Q_0 &= p_{01}^2 + p_{02}^2 + \cdots + p_{0k}^2 \\ &= \frac{1}{k^2} + \frac{1}{k^2} + \cdots + \frac{1}{k^2} \\ &= \frac{1}{k}. \end{aligned}$$

Consider another characteristic (h_1 say), which divides the system into k classes of unequal probability such that

$$p_j = p_{1j} = \frac{1}{k} + \epsilon_j; \quad j = 1, \ldots, k.$$

Since $\Sigma_{j=1}^{k} p_j = 1$ it can be inferred that $\Sigma_{j=1}^{k} \epsilon_j = 0$. Thus, for h_1,

$$\begin{aligned} Q = Q_1 &= p_{11}^2 + p_{12}^2 + \cdots + p_{1k}^2 \\ &= \Sigma_{j=1}^{k} \left(\frac{1}{k} + \epsilon_j \right)^2 \\ &= \Sigma_{j=1}^{k} \left(\frac{1}{k^2} + \frac{2\epsilon_j}{k} + \epsilon_j^2 \right) \\ &= \frac{1}{k} + \frac{2}{k} \Sigma_{j=1}^{k} \epsilon_j + \Sigma_{j=1}^{k} \epsilon_j^2 \\ &= \frac{1}{k} + \Sigma_{j=1}^{k} \epsilon_j^2 \quad \text{since} \quad \Sigma_{j=1}^{k} \epsilon = 0 \\ &\geq Q_0 \end{aligned}$$

since $\Sigma_{j=1}^{k} \epsilon_j^2$ is never negative (and equals zero if and only if $\epsilon_1 = \epsilon_2 = \cdots = \epsilon_k = 0$; i.e. if and only if $p_{0j} = p_{1j}$ $(j = 1, \ldots, k)$). Thus Q takes values between $1/k$ and 1 where k is the number of categories in the system. The probability of a match is minimised, and the discriminating power is maximised when the class probabilities are all equal. This is confirmation of a result that is intuitively reasonable, namely, that if a choice has to be made among several techniques as to which to implement, then techniques with greater variability (e.g. all categories are equally likely) should be preferred over those with lesser variability (e.g. only one category is possible so there no variation).

As an example of the application of this result, note that at the LDLR locus in Table 1.1 that

$k = 3$ and $1/k = 0.33$. Thus, Q cannot be lower than 0.33 for this locus and the discriminating power cannot be greater than 0.67. Notice also that the minimum value $(1/k)$ of Q decreases as k increases. Discriminating power increases as the number of categories into which an item can be classified increases, a result that is intuitively attractive.

The aforementioned calculations assume that the total population size (N say) of interest is very large and that for at least one p_j, p_j^2 is much greater than $1/N$. Failure of these assumptions can lead to misleading results as described by Jones (1972) with reference to the results of Tippett et al. (1968) on paint fragments; see Examples 3.2 and 3.3.

The experiment described by Tippett et al. (1968) compared two thousand samples of paint fragments, pairwise. For various reasons, the number of samples was reduced to 1969, all from different sources. These were examined by various tests and only two pairs of samples from different sources were found to be indistinguishable. The total number of pairs which can be picked out at random is $\frac{1}{2} \times 1969 \times 1968 = 1\ 937\ 496$. Two pairs of samples were found to agree with each other. The probability of picking a pair of fragments at random that are found to be indistinguishable is thus estimated empirically as $2/1\ 937\ 496 = 1/968\ 478$.

This result is used as an estimate of the probability of a match (Q). The method by which it was determined is extremely useful in situations

such as the one described by Tippett et al. (1968) in which frequency probabilities are unavailable and, indeed, for which a classification system has not been devised. The extremely low value (1/968 748) of Q demonstrates the high evidential value of the methods used by the authors. The conclusion from this experiment is that these methods are very good at differentiating between paints from different sources. Low values of Q were also obtained in work on head hairs (Gaudette and Keeping, 1974) and footwear (Groom and Lawton, 1987).

The equivalence of the theoretical and empirical approaches to the determination of Q can be verified numerically using the LDLR locus with genotypic frequencies as given in Table 1.1. Assume there is a sample of 1000 people with genotypic frequencies in the proportions in Table 1.1. All possible pairs of people in this sample are considered and their genotypes compared. There are $\frac{1}{2}(1000 \times 999) = 499\ 500$ ($= P$, say) different pairings. Of these pairings there are the following numbers of matches for

- genotype AA : $188 \times 187/2 = 175\ 78$,
- genotype BB : $321 \times 320/2 = 51\ 360$,
- genotype AB : $491 \times 490/2 = 120\ 295$.

There are, thus, $M = \{(188 \times 187) + (321 \times 320) + (491 \times 490)\}/2 = 189\ 233$ pairings of people who have the same genotype. The probability of a match, by this numerical method, is then

$M/P = 189{,}233/499{,}500 = 0.3788$. The probability of a match is $Q = p_1^2 + p_2^2 + p_3^2 = 0.188^2 + 0.321^2 + 0.491^2 = 0.3795$. The approximate equality of these two values is not a coincidence as a study of the construction of the ratio M/P shows.

The probability Q is sometimes called an *average probability* (Aitken and Robertson, 1987). An average probability provides a measure of the effectiveness of a particular type of transfer evidence at distinguishing between two randomly selected individuals (Thompson and Williams, 1991). In the context of blood stains, it is so-called because it is the average of the probabilities that an innocent person will be found to have an allele, which matches that of a crime stain. These probabilities are of the type considered in Example 3.1. For example, for locus TPOX in Table 3.2 if the crime stain were of allele 8, the probability an innocent suspect matches this is just the probability he has allele 8, namely, 0.554. A similar argument gives probabilities of 0.093, 0.054, 0.259 and 0.040 for matches between crime stain alleles and that of an innocent person of alleles 9, 10, 11, and 12, respectively. The average probability is the weighted average of these four probabilities, where the weights are the corresponding population proportions, namely 0.554, 0.093, 0.054, 0.259, and 0.040, respectively. The average probability is then just Q, given by $(0.554 \times 0.554) + (0.093 \times 0.093) + (0.054 \times 0.054) + (0.259 \times 0.259) + (0.040 \times 0.040) = 0.3872$.

3.5.3 Finite Samples

The relationship between the general result for a population, which is conceptually infinite in size, and a sample of finite size is explained by Jones (1972). Consider a test to distinguish between k classes $C_1, \ldots, C_k$. A sample of n individuals is taken from the relevant population. The number of individuals in each class is $c_1, c_2, \ldots, c_k$ with $\Sigma_{j=1}^{k} c_j = n$. An estimate of the probability that a randomly selected individual will be in class C_j is $\hat{p}_j = c_j/n,\ j = 1, 2, \ldots, k$.

There are $n(n-1)/2$ possible pairings of individuals. For any particular class j, the number of pairings of individuals within the class is $c_j(c_j - 1)/2,\ j = 1, 2, \ldots, k$. Thus, the overall proportion of pairings which result in a match is

$$\hat{Q} = \{\Sigma_{j=1}^{k} c_j(c_j - 1)\}/\{n(n-1)\}.$$

Then

$$\begin{aligned}
\hat{Q} &= (\Sigma_{j=1}^{k} c_j^2 - \Sigma_{j=1}^{k} c_j)/\{n(n-1)\} \\
&= (\Sigma_{j=1}^{k} c_j^2 - n)/(n^2 - n) \text{ since } \Sigma_{j=1}^{k} c_j = n, \\
&= \{\Sigma_{j=1}^{k} (c_j^2/n^2) - 1/n\}/(1 - 1/n) \\
&= (\Sigma_{j=1}^{k} \hat{p}_j^2 - 1/n)/(1 - 1/n). \qquad (3.3)
\end{aligned}$$

When the class frequencies are known, the probability of a match PM is given by Q (3.2). Result (3.3) gives an exact expression for the probability of a match for any given sample for all values of n and

$\{\hat{p}_j, j = 1, \ldots, k\}$. As n increases towards the population size, which is assumed to be so large that its reciprocal can be ignored in the calculation, it is to be expected that the observed sample probability will converge to the population probability. For n large it can be seen that $\hat{Q}$ tends to

$$\Sigma_{j=1}^{k} \hat{p}_j^2,$$

since $1/n$ becomes vanishingly small. As $\hat{p}_j$ tends to p_j so $\Sigma_{j=1}^{k} \hat{p}_j^2$ will tend to Q. However, as well as n being large, it is necessary for at least one of the $\hat{p}_j^2$ to be much greater than $1/n$ in order that $\Sigma \hat{p}_j^2 - 1/n \simeq \Sigma p_j^2$. This should be so for k not too close to n or for n very much larger than k; i.e. the number, n, of individuals should be much greater than the number of categories k.

Two examples (Jones, 1972) are given in which the probability of a match estimated by $\hat{Q}_1 = \Sigma \hat{p}_j^2$ is not a good approximation to the probability of a match estimated by $\hat{Q}$ from (3.3). The true probability of a match $Q = \Sigma p_j^2$ will not often be known exactly, except in situations like blood grouping systems where the sample sizes are extremely large and the $\{p_j\}$ are known accurately.

Example 3.2. If n is small, then $1/n$ is not very small compared with 1. Consider four playing cards, of which two are red (R_1, R_2: Category 1) and two are black (B_1, B_2: Category 2). Thus, $n = 4$, $c_1 = c_2 = 2$, $p_1 = \hat{p}_1 = p_2 = \hat{p}_2 = 1/2$ and $Q = \hat{Q}_1 = 1/4 + 1/4 = 1/2$. Note that $1/n = 1/4$

which is not very much less than 1. There are six possible pairings of cards $(R_1R_2, R_1B_1, R_1B_2, R_2B_1, R_2B_2, B_1B_2)$ of which two result in a match (R_1R_2, B_1B_2). Thus $\hat{Q} = 1/3$ which may be verified from $\hat{Q} = (\Sigma\hat{p}_j^2 - 1/n)/(1 - 1/n) = (1/2 - 1/4)/(1 - 1/4) = 1/3$. The failure of $1/n$ to be very small has led to a discrepancy between $\hat{Q}$ and $\Sigma_{j=1}^{k}\hat{p}_j^2 = \frac{1}{2}^2 + \frac{1}{2}^2 = \frac{1}{2}$.

Example 3.3. Consider the paper by Tippett et al. (1968) in which 1969 sets of paint fragments were considered ($n = 1969$) and there are very small values of p_j^2. Two pairs were found to be indistinguishable. Label these pairs as belonging to classes 1 and 2. The other paint fragments may be said to belong to 1965 different classes labelled $3, \ldots, 1967$, each with only one member so that $\hat{p}_1 = 2/1969,\ \hat{p}_2 = 2/1969,\ \hat{p}_3 = \cdots = \hat{p}_{1967} = 1/1969$. Then

$$\begin{aligned}\hat{Q}_1 &= \left(\frac{2}{1969}\right)^2 + \left(\frac{2}{1969}\right)^2 + \left(\frac{1}{1969}\right)^2 \\ &\quad + \cdots + \left(\frac{1}{1969}\right)^2 \\ &= \frac{1973}{1969^2} \\ &\simeq \frac{1}{1965},\end{aligned}$$

whereas

$$\hat{Q} = \left(\frac{1973}{1969^2} - \frac{1}{1969}\right) / \left(1 - \frac{1}{1969}\right)$$

$$= \frac{4}{1969 \times 1968}$$
$$= \frac{1}{968\ 748},$$

agreeing with the earlier result obtained by the authors. Here the approximate result $\hat{Q}_1$ is very inaccurate because no $\hat{p}_j^2$ is very much greater than $1/n$. In fact, the largest $\hat{p}_j^2$ is $(2/1969)^2$, which is smaller than $1/n = 1/1969$.

The use of the discriminating power for assessing the value of a comparative technique has been discussed in detail by Evett (1990).

3.5.4 Combination of Independent Systems

Consider Q from (3.2). This is the probability of finding a match between two individuals selected at random using a particular classification system. Suppose now that there are p independent systems with corresponding Q values $Q_1, \ldots, Q_p$. The probability of finding a pair that matches on all p tests is $\mathrm{PM}_p = \prod_{l=1}^{p} Q_l$. The probability of being able to distinguish between two individuals using these p tests is therefore

$$\mathrm{DP}_p = 1 - \prod_{l=1}^{p} Q_l.$$

Consider, for the sake of illustration, the following example for a comparison of the allelic

frequencies between New Zealand (NZ) and Swiss Caucasians, with frequency results for the TPOX and TH01 loci. The calculations for the discriminating power for the combination of the TPOX and TH01 systems are given later. The New Zealand data are given in Harbison et al. (2002). The Swiss data are given by courtesy of the Centre Universitaire de Médecine Légale (The University of Lausanne).

The allelic frequencies for the TPOX and TH01 systems for NZ and Swiss Caucasians are given in Tables 3.2 and 3.3.

The probability that two blood samples match on both criteria is

$$\begin{aligned} PM_2 &= Q_{TPOX} \times Q_{TH01} \\ &= 0.3872 \times 0.2305 = 0.0892 \text{ (Swiss)}, \\ &= 0.3780 \times 0.2251 = 0.0851 \text{ (NZ)}. \end{aligned}$$

Table 3.2 Allelic frequencies for TPOX locus for Swiss and NZ Caucasians and the probability Q_{TPOX} of a match.

Allele	Frequency	
	Swiss	NZ
8	0.554	0.529
9	0.093	0.082
10	0.054	0.063
11	0.259	0.294
12	0.040	0.032
Q_{TPOX}	0.3872	0.3780

Table 3.3 Allelic frequencies for TH01 locus for Swiss and NZ Caucasians and the probability Q_{TH01} of a match.

Allele	Frequency	
	Swiss	NZ
5	0.0	0.002
6	0.219	0.180
7	0.194	0.206
8	0.083	0.102
9	0.144	0.155
9.3	0.342	0.340
10	0.018	0.015
Q_{TH01}	0.2305	0.2251

The discriminating power is

$$\begin{aligned} \text{DP}_2 &= 0.9108 \text{ (Swiss)}, \\ &= 0.9149 \text{ (NZ)}. \end{aligned}$$

3.5.5 Correlated Attributes

Consider the hair example of Gaudette and Keeping (1974) in more detail. A similar argument holds for the paint example of Tippett et al. (1968). A series of 366 630 pairwise comparisons between hairs from different individuals were made. Nine pairs of hairs were found to be indistinguishable. These results were used to provide an estimate of the probability that a hair taken at random from one individual,

A, say, would be indistinguishable from a hair taken at random from another individual, *B*, say, namely, 9/366 630 or 1/40 737. It was then argued that if nine dissimilar hairs were independently chosen to represent the hairs on the scalp of individual *B*, the chance that a single hair for *A* is distinguishable from all nine of *B*'s may be taken as $[1-(1/40\ 737)]^9$, which is approximately $1-(1/4\ 500)$. The complementary probability, the probability that a single hair from *A* is indistinguishable from at least one of *B*'s hairs is 1/4 500. This probability provides, in some sense, a measure of the effectiveness of human hair comparison in forensic hair investigations.

There are various criticisms, though, which can be made regarding this approach, some details of which are in Aitken and Robertson (1987). Comments on the Gaudette–Keeping study are also made in Fienberg (1989). Criticisms on the methodology are presented in Barnett and Ogle (1982) and Miller (1987), with rebuttals in Gaudette (1982, 1999).

First note that the assumption of independence of the nine hairs used in the calculation is not an important one. The use of an inequality known as the *Bonferroni inequality* gives an upper bound for the probability investigated by the authors of 1/4 526 (Gaudette, 1982). The Bonferroni inequality states that the probability that at least one of several events occurs is never greater than the sum of the probabilities of the occurrences

of the individual events. Denote the events by $R_1, R_2, \ldots, R_n$, then the inequality states that

$$\Pr(\text{at least one of } R_1, \ldots, R_n \text{ occurs}) \leq \Pr(R_1) + \cdots + \Pr(R_n).$$

Gaudette and Keeping (1974) compared each of nine hairs, known to be from one source with one hair known to be from another source. The events $R_1, \ldots, R_9$, where $n = 9$ correspond to the inability to distinguish each of the nine hairs from the single hair. The probability of interest is the probability of at least one indistinguishable pair in these nine comparisons. This is the probability that at least one of $R_1, \ldots, R_9$ occurs. Using the Bonferroni inequality, it can be seen that this probability is never greater than the sum of the individual probabilities. From earlier, these individual probabilities are all equal to 1/40 737. The sum of the nine of them then equals 9/40 737, which is 1/4 526. This is very close to the figure of 1/4 500 quoted from the original experiment. Even if independence is not assumed there is very little change in the probability figure quoted as a measure of the value of the evidence.

An important criticism, though, is the following. One probability of interest to the court is the probability that a hair found at a crime scene belonged to a defendant. Other probabilities of interest to the court, the relevance of which will

be explained in more detail later in Section 5.3.1, are the probability of the evidence of the similarity of the hairs (crime and defendant) if they came from the same origin and the probability of the evidence of the similarity of the hairs if they came from different origins. Gaudette and Keeping (1974) provided an estimate of the probability that hairs 'selected at random' from two individuals are indistinguishable. This probability is an average probability (Section 3.5.2). It can be used as a broad guideline to indicate the effectiveness of hair identification in general. However, the use of the figure 1/4 500 as the value of the evidence in a particular case could be very misleading.

The average probability is the probability that two individuals chosen at random will be indistinguishable with respect to the trait under examination. However, in a particular investigation one sample is of known origin (in this case, the sample of hair from the PoI, which is the control sample), the other (the recovered sample) is not. If the PoI is not the criminal, someone else is and the correct probability of interest is the probability that a person chosen at random (which is how the criminal must be considered) from some population will have hair that is similar to that of the PoI – see Chapter 6 for a more detailed discussion.

Fienberg (1989) also makes the point that, 'even if we interpret 1/4 500 ... as the probability of a match given a comparison of a suspected hair with a sample from a different individual we

would still need the probability of a match given a sample from the same individual as well as *a priori* probabilities of "same" and "different".'

3.6 SIGNIFICANCE PROBABILITIES

Procedures for the evaluation of evidence in forensic science were changed dramatically by a paper by Dennis Lindley in 1977 (Lindley, 1977c). Previous to the publication of that paper a common procedure was a two-stage approach.

- Similarity: In a comparison of characteristics of evidence found at a crime scene and in the environment of a PoI, are the characteristics similar or dissimilar?
- Rarity: If the characteristics are dissimilar then the evidence is not considered any further, the evidence associated with the PoI is deemed not to be associated with the crime. If the characteristics are similar then the evidence associated with the PoI is deemed to be associated with the crime. The strength of the evidence under source level propositions (Section 5.3.1) is measured by the rarity of the characteristic; the more rare the characteristic, the stronger the association.

This description of the two-stage approach begs the questions of what is meant by similarity and what is meant by rarity.

In the 1960s and the 1970s attempts were made to evaluate evidence using the principles of significance tests and probabilities.

3.6.1 Calculation of Significance Probabilities

During the course of a crime a window has been broken. A PoI is apprehended soon afterwards and a fragment of glass found on his clothing. Denote this fragment by F. It is of interest to assess the uncertainty relating to whether the fragment came from the broken window. The assessment will be discussed for the moment in the context of an approach based on what are known as *significance probabilities*. Later, in Sections 3.7 and 3.8, two other approaches to this assessment problem, based on coincidence probabilities and likelihood ratios, respectively, will be discussed.

Let θ_0 be the value of the parameter representing the mean refractive index of the broken window. This is assumed constant. In Section 3.7 this assumption will be dropped and θ_0 will be replaced by a set of sample measurements of the refractive index from the broken window or by a distribution on the parameter θ in Chapter 7.

Let x be the refractive index for F. This measurement may be considered as an observation of a random variable X, there being variation in refractive index within a particular window. It is assumed that if F came from a window of mean

refractive index θ then X is such that

$$X \sim N(\theta, \sigma^2),$$

(Section A.3.2). The question at issue is whether F came from the window at the crime scene (and, by association, that the PoI was present at the crime scene). If this is so then θ will be equal to θ_0.

An argument based on significance probabilities is as follows. Suppose $\theta = \theta_0$. The further inference that F came from the window at the crime scene requires the assumption that the mean refractive index is unique to that window; this is not a particularly statistical assumption and is something that should perhaps be part of I, the background information.

The supposition that $\theta = \theta_0$ will be referred to as the *null hypothesis* and denoted H_0. Other names for such a hypothesis are *working hypothesis* or *status quo*. This nomenclature is not particularly appropriate here. It does not seem reasonable to start the analysis with the hypothesis that the PoI was at the scene of the crime. Nonetheless, this line of reasoning is pursued as the statistical ideas on which it is based are in very common usage.

Under the supposition that $\theta = \theta_0$, the deviation (in absolute terms, independent of sign) of x, the refractive index of F, from θ_0 would be expected to be small – just what is meant by small depends on σ, the standard deviation of X. The distribution of X has been taken to be Normal. The deviation of an observation x from the mean θ of the distribution is

measured in terms of the probability of observing a value for the random variable X as extreme as x.

If H_0 is true,

$$X \sim N(\theta_0, \sigma^2).$$

Let $Z = (X - \theta_0)/\sigma$. Then

$$Z \sim N(0, 1).$$

Also, $\Pr(|\ X\ | > x \mid \theta_0, \sigma) = \Pr(|\ Z\ | > z)$.

The probability $\Pr(|\ X\ | > x \mid \theta_0, \sigma)$ is the probability of what has been observed (x) or anything more extreme if H_0 ($\theta = \theta_0$) is true (and hence, as discussed above, that F came from the window at the crime scene). The phrase 'anything more extreme' is taken to mean anything more extreme in relationship to an implicit alternative hypothesis that if H_0 is not true then $\theta \neq \theta_0$. The distance of an observation x from a mean θ is measured in terms of the standard deviation σ. For example, if $\theta = 1.518\ 458$ and $\sigma = 4 \times 10^{-5}$, a value of x of $1.518\ 538$ is $(1.518\ 538 - 1.518\ 458)/4 \times 10^{-5}$, or 2.0, standard deviations from the mean. A value of the refractive index x that is more extreme than $1.518\ 538$ is one which is more than 2 standard deviations from the mean in *either* direction, i.e. a value of x which is greater than $1.518\ 538$ or less than $1.518\ 378$.

The probability of what is observed, or anything more extreme, calculated assuming the null hypothesis is true, is known as the *significance probability*. It may be thought to provide a measure

of compatibility of the data with the null hypothesis. It is conventionally denoted P. A small value of P casts doubt on the null hypothesis. In the example discussed here, a small value would cast doubt on the hypothesis that F came from the window at the crime scene sufficient to enable a decision to be taken to act as if the null hypothesis were false. However, it is not clear what value to take for 'small'. See Wasserstein and Lazar (2016) for a general statement by the American Statistical Association on P-values and Taroni et al. (2016) for a critical discussion on the (ab)use of P-values in forensic science.

Certain values of P have been used to provide values known as *significance levels* at which a scientist decides to act as if the null hypothesis is false.[4] Typical values are 0.10, 0.05, and 0.01. Thus, for example, a value x of the refractive index for which $P < 0.05$ would be said to be *significant at the 5% level*. If an approach to the evaluation of evidence based on significance levels is taken then the choice of what level of P to choose is obviously of crucial importance. It is helpful when deciding on a level for P to bear in mind the implications of the decision. The significance probability P is the probability of what is observed or anything more extreme if the null hypothesis is true. Suppose a significance level of 0.05 has been chosen. Then, by chance alone, on 5% of occasions on which this

[4]For an examination of the origins of 'statistical significance', see Shafer (2019).

test is conducted, a decision will be made to act as if the null hypothesis is false and a wrong decision will have been made.

Consider the glass example again. On 5% of occasions, using a significance level of 5%, in which such a test was conducted and in which F came from the window at the crime scene, it will be decided to act as if F did not come from the window. This decision will be wrong. Obviously there will be many other factors that affect the decision in any particular case. However, the principle remains. The use of this type of analysis gives rise to the probability of an error. It is a well-known error and is given the name *type 1 error* or *error of the first kind*.

The probability of a type 1 error can be reduced by reducing the significance level, e.g. to 0.01 or 0.001. However, this can only be done at the expense of increasing the probability of another type of error, known as a *type 2 error* or *error of the second kind*. A type 2 error is the error of failing to reject the null hypothesis when it is false. In the example, it will be the error of deciding that F did come from the window at the crime scene when in fact it did not. In general, if other factors, such as the number of fragments considered, remain constant, it is not possible to choose a significance level so as to decrease the probabilities of both type 1 and type 2 errors simultaneously.

Assume F came from the window at the scene of the crime. Then $\theta = \theta_0$. As a numerical illustration, let $\theta_0 = 1.518\ 458$ and $\sigma = 4 \times 10^{-5}$.

Table 3.4 Significance probabilities P for refractive index x of glass for mean $\theta_0 = 1.518\,458$ and standard deviation $\sigma = 4 \times 10^{-5}$ and decisions assuming a significance level of 5%.

x	$z = (x - \theta_0)/\sigma)$	$P = \Pr(\mid X \mid > x) = \Pr(\mid Z \mid > z)$	Decide to act as if F *did* or *did not* come from the crime window
1.518 500	1.05	0.29	Did
1.518 540	2.05	0.04	Did not
1.518 560	2.55	0.01	Did not

Illustrations of the calculations for the significance probability are given in Table 3.4.

Determination of the probability of a type 2 error requires knowledge of the value of θ if F did not come from the window at the crime scene. Normally such knowledge is not available and the probability cannot be determined. There may, occasionally, be circumstances in which the probability of a type 2 error can be determined. For example, this could happen if the defence were to identify another source for F. However, even if a probability cannot be determined, one has to be aware that such an error may be made.

Notice that the philosophy described here fits uncomfortably with the legal philosophy that a person is innocent until proven guilty. The null hypothesis in the example is that F came from the

window at the crime scene. Failure to reject this hypothesis is an implicit acceptance that the PoI was at the crime scene. Yet, it is the null hypothesis that is being tested. It is the one that has to be rejected or not. The calculation of the significance probability P is based on the assumption that the null hypothesis is true. Only if P is small (and it is values of 0.05, 0.01, and 0.001, which have been suggested as small) is it considered that the null hypothesis may be false. Evidence is required to show that the PoI has *not* been at the crime scene. The principle that one is innocent until proven guilty requires that evidence be produced to show that the PoI has been at the crime scene, not the opposite. The burden of proof has shifted from the prosecution to defence.

This point is also made by Gaudette (1999) in the context of hair examination. An analogy is made with a fire alarm. A type 1 error corresponds to the alarm not ringing when there is a fire. A type 2 error corresponds to the alarm ringing when there is not a fire. With a fire alarm a type 1 error is more serious error than a type 2 error. In forensic science a type 2 error is more serious than a type 1 error as it results in the false incrimination of an innocent person.

The interpretation of P has to be made carefully. Consider a result from Table 3.4: $\theta_0 = 1.518\ 458$, $\sigma = 4 \times 10^{-5}$, $x = 1.518\ 560$, and $P = 0.01$. This may be written more explicitly in the notation of

conditional probability as

$$\Pr(|\ X\ |> x \mid \theta = \theta_0, \sigma) = 0.01. \qquad (3.4)$$

It is difficult to relate this to the matter at issue: was the PoI at the crime scene? A small value would seem indicative of the falsity of the null hypothesis but it is not the probability that the null hypothesis is true. It is incorrect to use the value for P as the probability that the suspect was at the crime scene. The transposed probability

$$\Pr(\theta = \theta_0 \mid |\ X\ |> x, \sigma) \qquad (3.5)$$

would be much more useful but this is not what has been calculated. The relationship between (3.4) and (3.5) is similar to that between the probability of the evidence given the PoI is guilty and the probability the PoI is guilty given the evidence. The interpretation of the first of the last two probabilities as the second probability is the fallacy of the transposed conditional (Section 2.5.1). It is possible, however, to discuss the relationship between the significance probability and the probability that the PoI was at the crime scene, through the use of likelihood ratios.

3.6.2 Relationship to Likelihood Ratio

The relationship between significance probabilities and the likelihood ratio has been investigated by many authors. Early references include Good

(1956), Lindley (1957), and Edwards et al. (1963). The discussion here in the context of the refractive index of glass is based on Berger and Sellke (1987).

Consider two competing and complementary propositions:

H_p: the fragment of glass on the clothing of the PoI came from the window at the crime scene;

H_d: the fragment of glass on the clothing of the PoI did not come from the window at the crime scene.

Let p denote the probability that H_p is true, $\Pr(H_p)$, and $(1-p)$ the probability that H_d is true, $\Pr(H_d)$. If H_p is true then θ, the mean refractive index of the source of the fragment F on the suspect's clothing is θ_0. If H_d is true, then F is assumed to come from a window whose mean refractive index is not equal to θ_0.

Assume H_p is true. Denote the probability density function of X, the refractive index of F, by $f(x \mid \theta_0, H_p)$. In this context, this is a Normal density function. More details on probability density functions and on distributions are presented in Chapter 7 and in Appendix A.

Assume H_d is true. Denote the probability density function of X by $f(x \mid H_d)$. The mean θ of the refractive index of the source of the fragment on the suspect's clothing may also be thought of as a random variable which varies from window to window over some relevant population of windows. As such it also has a probability

density function; denote this by $f(\theta)$. If θ is known, the probability density function of x is given by $f(x \mid \theta)$. An extension of the law of total probability (Section 1.7.10) to continuous data with integration replacing summation may be used to give the expression

$$f(x \mid H_d) = \int f(x \mid \theta) f(\theta) d\theta.$$

The probability density function of X, independent of H_p and H_d, is then

$$f(x) = p f(x \mid \theta_0, H_p) + (1 - p) f(x \mid H_d).$$

Thus, the probability that H_p is true, given the measurement x is

$$\begin{aligned} \Pr(H_p \mid x) &= f(x \mid \theta_0, H_p)\, p / f(x) \\ &= \left\{ 1 + \frac{(1-p) f(x)}{p f(x \mid \theta_0, H_p)} \right\}^{-1}, \quad (3.6) \end{aligned}$$

an expression similar to one used in paternity cases (see Section 6.3.4).

The posterior odds in favour of H_p are then, using a version of (2.14),

$$\frac{\Pr(H_p \mid x)}{1 - \Pr(H_p \mid x)} = \frac{p}{1-p} \frac{f(x \mid \theta_0, H_p)}{f(x \mid H_d)}$$

where $\{p/(1-p)\}$ is the prior odds in favour of H_p and $\{f(x \mid \theta_0, H_p)/f(x \mid H_d)\}$ is the likelihood ratio V of (2.15).

As an illustration of the calculation of the likelihood ratio, assume that if θ is not equal to

θ_0 it is a random variable, which has a Normal distribution with mean θ_0 and variance τ^2, where, typically, $\tau^2 \gg \sigma^2$, i.e, the variation in refractive index between windows is much greater than the variation in refractive index within windows. Then

$$f(x \mid H_d) = \int f(x \mid \theta) f(\theta)\, d\theta$$

and so

$$(X \mid H_d) \sim N(\theta_0, \sigma^2 + \tau^2).$$

(See Section 7.4.1 for a fuller derivation of this result.) With $\tau^2 \gg \sigma^2$, the distribution of $(X \mid H_d)$ is approximately $N(\theta_0, \tau^2)$. The likelihood ratio is, thus,

$$\begin{aligned} \mathrm{V} &= \frac{f(x \mid \theta_0, H_p)}{f(x \mid H_d)} \\ &= \frac{(2\pi\sigma^2)^{-1/2} \exp\{-(x-\theta_0)^2/2\sigma^2\}}{(2\pi\tau^2)^{-1/2} \exp\{-(x-\theta_0)^2/2\tau^2\}}. \end{aligned}$$

Consider $\tau = 100\sigma$. Let $z^2 = (x-\theta_0)^2/\sigma^2$ be the square of the standardised distance between the observation x and the mean specified by the null hypothesis, θ_0. Then

$$\begin{aligned} V &= 100 \exp\left(-\frac{z^2}{2} + \frac{z^2}{2 \times 10^4}\right) \\ &\simeq 100 \exp(-z^2/2). \end{aligned}$$

For example, consider $x = 1.518\ 540$, $\theta_0 = 1.518\ 458$, $\sigma = 4 \times 10^{-5}$, $\tau = 4 \times 10^{-3}$. Then

$z^2 = 2.05^2$ and $P = 0.04$, as before (see Table 3.4) which, at the 5% level, would lead to rejection of the hypothesis that the fragment of glass came from the window at the scene of the crime. However,

$$\begin{aligned} V &= 100 \exp(-2.05^2/2) \\ &= 12.2, \end{aligned}$$

a value for V that, on the verbal scale in Table 2.8, represents moderate evidence to support H_p against H_d. Such an apparent contradiction between the two approaches is not a new idea and has been graced by the name of Lindley's paradox(see, for example, Good (1956), Lindley (1957), Edwards et al. (1963), and Lindley (1980) for a reference by Lindley to 'Jeffreys' paradox'.)

Suppose that several, n, say, fragments of glass were found on the suspect's clothing rather than just one. Let $\bar{x}$ denote the mean of these fragments. Then,

$$(\bar{X} \mid \theta) \sim N(\theta, \sigma^2/n).$$

If H_d is true

$$\bar{X} \sim N(\theta_0, \tau^2 + \sigma^2/n).$$

The likelihood ratio is

$$V = \frac{(2\pi\sigma^2/n)^{-1/2} \exp\{-n(\bar{x} - \theta_0)^2/2\sigma^2\}}{\{2\pi(\tau^2 + \sigma^2/n)\}^{-1/2} \exp[-(\bar{x} - \theta_0)^2/\{2(\tau^2 + \sigma^2/n)\}]}$$

$$\simeq \frac{\tau\sqrt{n}}{\sigma}\exp\left[-\frac{(\bar{x}-\theta_0)^2}{2}\left\{\frac{n}{\sigma^2}-\frac{1}{\tau^2}\right\}\right]$$
$$\simeq 100\sqrt{n}\exp\left\{-\frac{n(\bar{x}-\theta_0)^2}{2\sigma^2}\right\}.$$

The square of the standardised distance, z_n, between $\bar{x}$ and θ_0 is

$$z_n^2 = n(\bar{x}-\theta_0)^2/\sigma^2$$

and thus

$$V = 100\sqrt{n}\exp(-z_n^2/2), \qquad (3.7)$$

a value that increases in direct proportion to the square root of the sample size. Suppose $z_n = 2$, a value that is significant at the 5% level in a test of the hypothesis $\theta = \theta_0$ against the alternative hypothesis $\theta \neq \theta_0$. The value of V for various values of n is given in Table 3.5. In each case, a result that is significant at the 5% level (and hence suggests rejection of the null hypothesis that $\theta = \theta_0$) has a likelihood ratio that lends support to the hypothesis that the fragments of glass found on the suspect came from the window at the crime scene.

3.6.3 Combination of Significance Probabilities

Significance probabilities also combine in different way from the probabilities of events. From the

Table 3.5 Variation in the likelihood ratio V, as given by (3.7), with sample size n, for a standardised distance $z_n = 2$, a result which is significant at the 5% level.

n	1	5	10	20
V	14	30	43	61

third law of probability (1.8), the product of two probabilities, for dependent or independent events, will never be greater than either of the individual components of the product. Let A and B be the two events. Then

$$\begin{aligned}\Pr(AB) &= \Pr(A)\Pr(B \mid A) \leq \Pr(A)\\ &= \Pr(B)\Pr(A \mid B) \leq \Pr(B)\end{aligned}$$

with equality in the first case if and only if $\Pr(B \mid A) = 1$ and in the second case if and only if $\Pr(A \mid B) = 1$. Note also that $\Pr(AB) > \Pr(A)\Pr(B)$ if $\Pr(B \mid A) > \Pr(B)$ or $\Pr(A \mid B) > \Pr(A)$.

However, it is possible, for characteristics that are dependent, that the significance probability of the joint observation may be greater than either of the individual significance probabilities.

Suppose that as well as the refractive index of the glass in the current example that the density has also been measured. The suffices 1 and 2 are introduced to denote refractive index and density, respectively. Then $\mathbf{x} = (x_1, x_2)^T$ is a vector that denotes the measurements of the two characteristics of the glass found on the PoI's clothing and

$\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ is a vector that denotes the mean values of refractive index and density of glass from the window at the crime scene. More details on vectors are given in Appendix B.

Let $\theta_1 = 1.518\,458$ (the original θ_0 earlier), $x_1 = 1.518\ 540$ and $\sigma_1 = 4 \times 10^{-5}$. Then, the significance probability (P_1, say) for the refractive index of F is 0.04 (see Table 3.4). Suppose $\theta_2 = 2.515\,\text{g cm}^{-3}$ with standard deviation $\sigma_2 = 3 \times 10^{-4}\text{g cm}^{-3}$ and let $x_2 = 2.515\ 615\text{g cm}^{-3}$ be the density measured on F. The standardised statistic, z_2 say, is then

$$z_2 = (2.515\ 615 - 2.515)/0.0003 = 2.05$$

and the significance probability (P_2, say) for the density measurement is also 0.04.

The product of P_1 and P_2 is 0.0016. However, this is not the overall significance probability. The correlation between the refractive index and the density has to be considered.

Let the correlation coefficient, ρ, say, between refractive index and density be 0.93 (Dabbs and Pearson, 1970). This value is used here to illustrate the line of reasoning. Assume that the joint probability density function of refractive index and density is bivariate Normal, Section A.3.9, with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$. For bivariate Normal data $\mathbf{x}$ it can be shown that the statistic U, given by

$$U = (\mathbf{x} - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\theta})$$

has a chi-squared distribution with 2 degrees of freedom:

$$U \sim \chi^2_2,$$

see Mardia et al. (1979, p. 39) and Section A.3.5. Values of the χ^2 distribution may be found from statistical software.

The overall significance probability for the two characteristics can be determined by calculating U and referring the answer to the χ^2 distribution. The covariance matrix Σ is given by (A.29)

$$\begin{aligned}\Sigma &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \\ &= \begin{pmatrix} (4\times 10^{-5})^2 & 1.116\times 10^{-8} \\ 1.116\times 10^{-8} & (3\times 10^{-4})^2 \end{pmatrix} \\ &= \begin{pmatrix} 1.6\times 10^{-9} & 1.116\times 10^{-8} \\ 1.116\times 10^{-8} & 9\times 10^{-8} \end{pmatrix}.\end{aligned}$$

The deviation of the observation $\mathbf{x}$ from the mean $\boldsymbol{\theta}$ is $(\mathbf{x}-\boldsymbol{\theta})^T = (8.2\times 10^{-5},\ 6.15\times 10^{-4})$. Some rather tedious arithmetic gives the result that $(\mathbf{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})$ equals 4.204, a result that has a significance probability $P = 0.1225$. Each individual characteristic is conventionally significant at the 5% level yet together the result is not significant at the 10% level.

Considerable care has to be taken in the interpretation of significance probabilities. As a measure of the value of the evidence their interpretation is difficult.

3.7 COINCIDENCE PROBABILITIES

3.7.1 Introduction

One of the criticisms that was levelled at the probabilistic evidence in the *Collins'* case (Section 3.4) was the lack of justification for the relative frequency figures that were quoted in court. The works of Tippett et al. (1968) and Gaudette and Keeping (1974) were early attempts to collect data. The lack of data relating to the distribution of measurements on a characteristic of interest is a problem that still exists and is perceived as a major one in forensic science. A detailed discussion on this topic – related to the analysis of the *R. v. T.* (2010) case – has been published by Berger et al. (2011), Aitken (2012), Biedermann et al. (2012b,c), Bodziak (2012a,b), Ligertwood and Edmond (2012a,b), Nordgaard and Rasmusson (2012a,b), Sjerps and Berger (2012), Hamer (2012), Thompson (2012), and Berger and Sjerps (2012).

If measurements of certain characteristics (such as glass fragments and their refractive indices) of evidence left at the scene of a crime are found to be similar to evidence (measurements of the same characteristics) found on a PoI, then the forensic scientist wants to know

(a) how similar are they, and

(b) is this similarity one that exists between rare characteristics or common characteristics?

In certain cases, such data do exist, which can help to answer these questions.

The most obvious example of this is DNA profiling. There are various loci and the relative population frequencies of the various categories for each of the loci are well tabulated for different populations. Thus if a blood stain of a particular profile is found at the scene of a crime there are several possibilities:

(a) it came from the victim,

(b) it came from the criminal,

(c) it came from some third party.

If the first and third possibilities can be eliminated and the profile is only found in $x\%$ of the population, then the implication is that the criminal belongs to that $x\%$ of the population; i.e. they belong to a subset of the population with $Nx/100$ members, where N is the total population size and $Nx/100$ is of necessity no less than 1.[5]

At its simplest this implies that, *all other things being equal*, a person with this allele has probability $100/Nx$ of being the criminal. This is a consequence of the defender attorney's fallacy explained in Section 2.5.4.

Another example in which data exist, though not to the extent to which they do for DNA profiling, is the refractive index of glass particles. Data relating to the refractive index have been collected over several years by different laboratories to be

[5] If $Nx/100 < 1$ refer to Section 2.5.5.

able to assess the relative frequency of glass fragments found on different locations (e.g. clothing, footwear, hair combings, and pocket). A full list up to 2000 is provided by Curran et al. (2000).

From a statistical point of view, this approach may lead to a biased sample since only glass fragments that have been connected with crimes will be collected. However, as an example of the recognition of this problem, a survey of glass found on footpaths has been conducted (Walsh and Buckleton, 1986) to obtain data unrelated to crimes.

An approach based on probabilities, known as *coincidence probabilities*, with relation to the refractive index of glass was developed by Evett and Lambert in a series of papers (Evett, 1977, 1978; Evett and Lambert, 1982, 1984, 1985), using data from Dabbs and Pearson (1970).

As in Section 3.6 two questions were asked.

(1) Are the control and recovered fragments similar in some sense?

(2) If they are similar are the characteristics rare or common?

A very simple – and trivial – example of this is the following. An eye-witness to a crime says that they saw a person running from the scene of the crime and that this person had two arms. A PoI is found who also has two arms. The PoI is certainly similar to the person running away from the crime but the characteristic – two arms – is so

common as not to be worth mentioning. The elementary nature of the example may be extended by noting that if the eyewitness said the criminal was a tall man then any women or short men, despite having two arms, would be eliminated. In contrast, suppose that the eye-witness says that the person running away had only one arm and a PoI is found with only one arm. This similarity is much stronger evidence that the PoI was at the scene of the crime since the characteristic – one arm – is rare.

A more complicated – and considerably less trivial – application relates to the interpretation of evidence on the refractive index of glass. The coincidence method to be explained is capable of development in this context because of the existence of data for the distribution of measurements of the refractive index of glass. Development is not possible in the examples mentioned earlier of paint fragments and hair comparisons because appropriate data do not exist. A crime is committed and fragments of glass from a window are found at the scene of the crime; these are the control fragments. A PoI is found and they have fragments of glass on their person; these are the recovered (or transferred-particle) fragments. The interpretation of the data is treated in two stages corresponding to the two questions asked earlier. First, the refractive index measurements are compared using a statistical criterion that takes account of between and within-window variation. Secondly, if the two sets of measurements are

found to be similar, then the significance of the result is assessed by referring to suitable data.

Various assumptions are needed for these two stages to be applied. Those adopted here are the following.

(a) The refractive index measurements on glass fragments from a broken window have a Normal probability distribution centred on a mean value $\theta = \theta_0$ which is characteristic of that window and with a variance σ^2.

(b) The mean θ varies from window to window and in the population of windows θ has its own probability distribution of, as yet, unspecified form.

(c) The variance σ^2 is the same for all windows and it is known.

(d) All the transferred particle fragments are assumed to have come from the same source. A criterion based on the range of the measurements can be devised to check this, e.g. Evett (1978).

(e) The transferred particle fragments are all window glass.

3.7.2 Comparison Stage

The comparison is made by considering the difference, suitably scaled, in the mean of the source fragment measurements and the mean of the recovered fragment measurements. The test

statistic is

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma(n^{-1} + m^{-1})^{1/2}}$$

where $\bar{X}$ is the mean of m measurements on control fragments and $\bar{Y}$ is the mean of n measurements on recovered fragments. It is assumed that Z has a standard $N(0, 1)$ distribution. Thus if $|Z| > 1.96$ it is concluded that the recovered fragments are not similar to the source fragments and if $|Z| < 1.96$ it is concluded that the recovered fragments are similar to the control fragments. The value 1.96 is chosen so that the probability of a type 1 error (deciding wrongly that the recovered fragments have a different origin from the control fragments) is 0.05. Other possible values may be chosen such as 1.64, 2.33, and 2.58, which have type 1 error probabilities of 0.10, 0.02, and 0.01, respectively.

This statistic and the associated test provide an answer to question (1).

3.7.3 Significance Stage

If it is concluded that the recovered and control fragments are not similar, the matter ends there. If it is concluded that they are similar, then there are two possibilities.

(a) The recovered and control fragments came from the same source.

(b) The recovered and control fragments did not come from the same source.

For the assessment of significance the probability of a coincidence (known as a *coincidence probability*) is estimated. This is defined as 'the probability that a set of n fragments taken at random from some window selected at random from the population of windows would be found to be similar to a control window with mean refractive index $\bar{X}$'. It is denoted $C(\bar{X})$.

Compare this definition of a probability of coincidence with that given in Section 3.5. There the recovered and control fragments were both taken to be random samples from some underlying population and the probability estimated was that of two random samples having similar characteristics. Here the mean $\bar{X}$ of a sample of fragments from the control window is taken to be fixed and the concern is with the estimation of the probability, $C(\bar{X})$, say, that one sample, namely, the recovered fragments, will be found to be similar to this control window. Any variability in the value of $\bar{X}$ is ignored; methods of evaluation which account for this variability are discussed in Chapter 7.

Note that there are two levels of variation which have been considered. Further details of how to consider these two levels is described in Chapter 7. First there is the probability that a window selected at random from a population of windows will have a mean refractive index in a certain interval, $(u, u + du)$ say; denote this probability by $p(u)$. In practice for this method the data used to estimate $p(u)$ are represented as a

histogram and the probability distribution is taken as a discrete distribution over the categories forming the histogram, e.g. $\{p(u_1), \ldots, p(u_k)\}$ where there are k categories. Secondly, the probability is required that n fragments selected at random from a window of mean value u would prove to be similar to fragments from a source window of mean $\bar{X}$; denote this probability by $S_{\bar{X}}(u)$. It is then possible to express $C(\bar{X})$ as a function of $p(u)$ and $S_{\bar{X}}(u)$, namely:

$$C(\bar{X}) = \sum_{i=1}^{k} p(u_i) S_{\bar{X}}(u_i). \tag{3.8}$$

The derivation of this is given in Evett (1977).

Let $\{x_1, \ldots, x_m\}$, with mean $\bar{x}$, denote the source measurements of refractive index from fragments from a window W broken at the scene of a crime; with

$$\bar{X} \sim N(w, \sigma^2/m),$$

where w is the mean refractive index of window W, see Section A.3.2.

Let $\{y_1, \ldots, y_n\}$, with mean $\bar{y}$, denote the transferred particle measurements of refractive index from fragments T found on a suspect's clothing. If the fragments T have come from W then

$$\bar{Y} \sim N(w, \sigma^2/n).$$

Since the two sets of measurements are independent,

$$(\bar{X} - \bar{Y}) \sim N(0, \sigma^2(1/m + 1/n)),$$

and, hence,

$$\frac{\bar{X} - \bar{Y}}{\sigma(1/m + 1/m)^{1/2}} \sim N(0, 1).$$

Denote this statistic by Z. Thus, it can be said that if the fragments T came from W and if the distributional assumptions are correct, the probability that Z has a value greater than 1.96 in absolute terms; i.e.

$$| Z | > 1.96,$$

is 0.05 (Table A.5). Such a result may be thought unlikely under the original assumption that the fragments T have come from W. Hence the original assumption may be questioned. It may be decided to *act as if* the fragments T did not come from W. Evett (1977) showed that with this decision rule at the comparison stage, then the probability $S_{\bar{X}}(u)$ at the significance stage is given by

$$\begin{aligned} S_{\bar{X}}(u) = \Phi\{(\bar{X} - u)m^{1/2}/\sigma + 1.96(1 + m/n)^{1/2}\} \\ -\Phi\{(\bar{X} - u)m^{1/2}/\sigma - 1.96(1 + m/n)^{1/2}\}, \end{aligned} \tag{3.9}$$

(Section A.3.2) from which the coincidence probability $C(\bar{X})$ may be evaluated in any particular case. Some results of the application of this method in comparison with others are given in Section 7.5.3.2 and Table 7.12.

3.8 LIKELIHOOD RATIO

The likelihood ratio is discussed at considerable length in later chapters. The likelihood ratio has been introduced in Section 2.4.1 as the ratio $\Pr(E \mid H_p, I)/\Pr(E \mid H_d, I)$ (2.15), which converts prior odds in favour of the prosecution proposition (H_p) into posterior odds in favour of the prosecution proposition, given evidence E. Background information is denoted I.

From a historical point of view, remember that, in the Dreyfus case discussed in Section 3.2, Poincaré and his colleagues supported such an approach and proposed the use of the likelihood ratio. As Good (1979) points out the idea for the use of the likelihood ratio to evaluate evidence pre-dates the work of Poincaré and his colleagues. In an earlier article C.S. Peirce (1878) introduced the term 'weight of evidence' for the logarithm of the likelihood ratio (Section 2.3.4). Also

> *an effect may be the product of either cause A or a cause B. The effect has already been observed. One wants to know the probability that it is the result of cause A; this is the* a posteriori *probability. But I am not able to calculate in advance the* a priori *probability for the cause producing the effect. I want to speak of the probability of this eventuality, for one who has never before observed the result. (Poincaré, 1992, p. 229)*

However, as supported by Darboux et al. (1908),

> *since it is absolutely impossible for us [the experts] to know the* a priori *probability, we cannot say: this coincidence proves that the ratio of the forgery's probability to the inverse probability is a real value. We can only say: following the observation of this coincidence, this ratio becomes X times greater than before the observation. (p. 504)*

For more information on this statistical argument and an example of application to shoe prints, see Taroni et al. (1998).

Similarly, de Finetti (1930) expressed the same view:

> *Il calcolo delle probabilità è la logica del probabile. Come la logica formale insegna a dedurre la verità o la falsità di certe conseguenze della verità o falsità di certe premesse, così il calcolo delle probabilità insegna a dedurre la maggiore o minore verosimiglianza o probabilità di certe conseguenze dalla maggiore o minore verosimiglianza o probabilità di certe premesse. (p. 259)*

which can be translated as follows:

> *Probability calculus is the logic of the probable. As logic teaches the deduction of the truth or falseness of certain consequences from the truth or falseness of certain assumptions, so probability calculus teaches the deduction of the major or minor likelihood, or probability, of certain consequences from the major or minor likelihood, or probability, of certain assumptions.*

Olkin (1958) proposed an evaluation of the identification problem in terms of the likelihood ratio statistic, which – he said – is the ratio of the probability of the characteristics under the assumption

of the identity to the probability of the characteristics under the assumption of non-identity.

There are also some comments on the use of the likelihood ratio to assess the value of the scientific evidence in forensic science journals during the sixties as presented in Section 1.7.4. Kirk and Kingston were the main authors giving clear examples of this idea. For example, Kingston and Kirk (1964) reported that 'a general concept of probabilities will be examined to see what relationship it might bear to the interpretative areas of criminalistics' (p. 514). They presented an example of the likelihood ratio calculation in a glass fragments scenario. Kingston (1965b) extended the use of the likelihood ratio by presenting the use of the Bayesian approach and showing its relationship with the Essen-Möller formula[6] for the probability of paternity. More on the probability of paternity is presented in Section 6.3.4.

An interesting and somewhat prophetic comment was made by Parker and Holford (1968), in which the comparison problem was treated as in two stages (similarity and discrimination) rather than in one, as an analysis using the likelihood ratio does. The two-stage approach follows, they said, the traditions of forensic scientists.

[6]Essen-Möller – by applying the Bayesian computation – adopted the formula to compute the probability of paternity (a posterior probability) which includes both the likelihood ratio called *Paternity Index* and equal prior probability (in normal triplet cases).

Interestingly, they then went on to make the following remark (Parker and Holford, 1968):

> *We could therefore set up an index R whose numerator is the likelihood that the crime hair comes from the suspect and whose denominator is the likelihood that it comes from the population at large. In ordinary language one would then assert that it was R times more likely for the hair to have come from the suspect than from someone drawn at random from the population. But a statement of this nature is of rather limited use to forensic scientists, by-passing as it does the similarity question altogether. (p. 238)*

There is one flaw in this piece and that is that the authors have committed what is known as a 'source probability error' (see Section 2.5.2).

The authors may perhaps have committed this fallacy, but incidentally – and strictly speaking – their wording is technically correct because the words 'likelihood' and 'probability' are not synonyms: the probability of the evidence conditional on H is sometimes defined as the likelihood of H given E. Notationally

$$\Pr(E|H) = L(H; E).$$

Hence the term 'likelihood ratio is given to the ratio $\Pr(E \mid H_p)/\Pr(E \mid H_d)$ Thus when Parker and Holford (1968) say the numerator is the likelihood 'that the crime hair comes from', this is technically correct as it refers to the probability of the evidence given the proposition.

Note that often the expressions 'it is likely' and 'it is probable' are used interchangeably. For

example, it is often said as a verbal summary of the likelihood ratio that evidence is X times more likely if the prosecution proposition is true than if the defence proposition is true. However, the same cannot be said about likelihood and probability. The likelihood of H, given E and the probability of H conditional on E are different concepts. An illustrative example is presented in Taroni et al. (2014a):

> *As an illustration of the importance of the difference between likelihoods and probabilities, consider the following example. Mr Jones has been seen by an eyewitness running away from the house where a crime has been committed, at approximately the time of the crime. Let* E[7] *be the proposition 'Mr Jones was running away from the scene of the crime at the time when it was committed' and* H *be the proposition 'Mr Jones committed the crime'. It is reasonable to believe that the likelihood* $\Pr(E \mid H, I)$ *is high, but it is not necessarily the case that* $\Pr(H \mid E, I)$ *is high as well. If Mr Jones actually committed the crime, it is expected that he would want to hasten away from the scene of the crime. The hypothesis of culpability provides a good explanation for the evidence. However, the fact that he was running away from the house does not, by itself, make it very probable that he committed the crime. There are many other possible explanations of that fact. (p. 9)*

The distinction between the two terms is crucial, because likelihood and probabilities do not have the same properties. For example, the sum of likelihoods can be greater than 1 (e.g. both

[7] Note that in the original text letters B and A were used instead of E and H. The change has been made for notational consistency with the remainder of the text here.

numerator and denominator likelihoods can be 1), and the likelihood is fixing the evidence and varying the hypotheses.

The confusion of the likelihood of hypothesis H, given evidence E, with the probability of the same hypothesis, conditional on the same evidence, is the 'fallacy of the transposed conditional' (see Section 2.5.1). A discussion is also presented in Biedermann and Vuille (2018).

Parker and Holford (1968) give an opinion on the source of the hair that is not within the remit of the forensic scientist. The ratio to which the scientist can testify is, in this context, 'the evidence is R times more likely if the hair had come from the suspect than if it had come someone other from the population', where 'likely' is used as a synonym for 'probable'. Thus whilst 'likelihood' and 'probability' are not synonyms, 'likely' and 'probable' are. It will be explained later (see Chapters 5 to 7) how the use of the likelihood ratio, which is all that Parker and Holford's index R is, has rather more than 'limited use' and how its use considers similarity in a natural way.

The importance of the likelihood ratio for evidence evaluation has been recognised by many writers, e.g. Jeffrey (1975), Lempert (1977), and Robertson and Vignaux (1992):

> *Bayesianism does not take the task of scientific methodology to be that of establishing the truth of scientific hypotheses, but to be that of confirming or disconfirming them to degrees which reflect the overall effect of the*

> *available evidence – positive, negative, or neutral, as the case may be. (Jeffrey, 1975, p. 104),*

> *[E]vidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue. (Lempert, 1977, p. 1026)*

and

> *Bayes' rule*[8] *tells us that we then take those prior odds and multiply them by the likelihood ratio of the blood/DNA evidence in order to arrive at the posterior odds in favour of the defendant's paternity. The Court then has to consider whether those odds meet the required standard of proof. Thus the expert should say 'however likely you think it is that the defendant is the father on the basis of the other evidence, my evidence multiplies the odds X times' (Robertson and Vignaux, 1992, p. 316).*

[8]Bayes' theorem.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

4

Bayesian Inference

4.1 INTRODUCTION

Much of what will be discussed in this chapter concerns Bayesian estimation of the values of parameters characterising a probability distribution. A fundamental tenet of the Bayesian paradigm states that uncertainty of the knowledge of the value of a parameter is modelled with a probability distribution.

It is possible to make statements about the value of a parameter in the form of a probability distribution known as a *prior distribution* that summarises the available knowledge on the values of the parameter before data become available. Bayesian inference can incorporate subjective information about a problem into the analysis. One of the criticisms of the Bayesian approach to evidence evaluation (and also to other areas of statistical analysis) is the loss of objectivity that results from the use of the analyst's subjective

information. A recurrent problem is the choice of an appropriate prior distribution with a shape to reflect appropriately an analyst's beliefs. From a practical point of view, there is often the tendency to seek criteria for the determination of a so-called ignorance (or objective) prior distribution, for example, the uniform distribution, upon which many observers might agree as that which has minimal subjective information. In this way, an expert can pursue the idea of not committing themselves to a subjective choice that is likely to be perceived as questionable. However, it might be objected that the choice of a uniform prior distribution, for example, could itself be subject to criticism as a subjective choice and one that does not necessarily reflect an absence of opinion (Taroni et al., 2010). It must be added that there are often practical circumstances where it is difficult to accept the idea that a practitioner has no idea about the possible values for a parameter of interest, and that any value can therefore be reasonably treated as equally likely as any other. Forensic scientists often have access to considerable sources of information (such as literature and past experience) that could be explored to propose informed subjective distributions. The elicitation of a prior distribution from available information will be briefly addressed in this chapter (Section 4.3.1 with reference to a proportion and Section 4.5 with reference to a Normal mean), but

readers can refer to e.g. O'Hagan et al. (2006) for a wide treatment of this topic.

The discussion of the Bayesian inference in Chapter 2 is to discrete events. This chapter extends these ideas to continuous variables, including the inference about a proportion (Section 4.2), the problem of sampling (Sections 4.3 and 4.4), inference about a mean (Section 4.5), quantity estimation (Section 4.6), and decision analysis (Section 4.7). There will be further reference to Bayesian ideas in Chapter 7 and an example in paternity in Section 6.3.3. A general introduction to Bayesian ideas is given in Berry (1996), Antelman (1997), Taroni et al. (2010), Lee (2012), and Bolstad and Curran (2017).

Denote the prior parameter of interest by θ, which may be vector-valued, and the prior distribution by $f(\theta)$. This distribution has itself one or more parameters (known as prior parameters or hyperparameters) by which it is characterised. Data x are modelled by the likelihood. The likelihood involves x and θ. It is proportional to a probability density function (for continuous x) and to a probability mass function (for discrete x). Denote the likelihood by $L(\theta \mid x)$ to emphasise that it is of interest as a function of θ, conditional on the value of x. A general description of the inversion of probabilities, as given by Bayes' theorem, is presented in Section 2.2 for discrete events. The continuous version of Bayes' theorem enables

the combination of the likelihood $L(\theta \mid x)$ and the prior $f(\theta)$ to form a posterior density function $f(\theta \mid x)$ as

$$f(\theta \mid x) = \frac{L(\theta \mid x)\, f(\theta)}{f(x)}, \tag{4.1}$$

where $f(x)$ is the marginal probability density function for x, unconditional on θ. It can be determined as

$$f(x) = \int_{\Theta} f(x \mid \theta) f(\theta) d\theta.$$

Note that, whenever θ is discrete, the integral is replaced by a sum over all $\theta \in \Theta$, the parameter space. Bayes' theorem allows a sequential update of the posterior density as new data become available. The *posterior distribution* of θ given data x becomes the new prior distribution once new information becomes available. Bayes' theorem describes the process of learning from experience showing how beliefs about θ can be continually modified.

In the Bayesian paradigm, the data are taken as fixed and known. Uncertainty resides in the parameters. A parameter θ is not considered as fixed and unknown (as in the frequentist point of view) but as a random variable. The available information about θ and one's uncertainty about it is summarised in a probability density function $f(\theta)$. Bayesians do not question the existence of a fixed unknown constant. Examples of such fixed constants include the speed of light and the

Hubble constant. What Bayesians do is represent and summarise all the information available on θ, the estimate of the fixed constant, without questioning the existence of the constant.

The (posterior) distribution on θ summarises the available information and the distribution is updated as more data are acquired. The distribution describes the uncertainty that is associated with the data and the scientists' knowledge. It is assumed there is no physical variability associated with the constant (Robert, 2013).

The data modify the prior distribution via the likelihood to provide the posterior distribution. The Bayesian inference about a parameter of interest is therefore described as the modification of prior uncertainty about its unknown value in the light of evidence. There is a relation between the prior distribution, which may arise from elicitation, and posterior inference. One must be aware that different prior opinions, or a different model for the translation of the prior state of knowledge may affect Bayesian estimates and this may represent a source of concern. It may be that a different prior distribution, close to the original prior, could be equally acceptable, but it could also be that different prior specifications give rise to different posterior distributions (and decisions as discussed in Section 4.7). There are in fact instances where the amount of available data is considerable, and the posterior distribution depends more and more on the data and less and less on prior specifications. In such cases the

posterior distribution is said to be 'dominated by the likelihood'. No matter how different the prior specifications might be, the corresponding posterior distributions will tend to be close. However, in many practical situations this is not the case. Available data might be very poor, possibly characterised by large variability, with the immediate consequence that personal knowledge could be relevant in determining the posterior distribution. Practitioners should perform a sensitivity analysis to explore the robustness of posterior inference to different specifications of the prior distribution. An example will be provided in Section 4.5.1.

The posterior distribution thus encapsulates all the available knowledge about θ. However, as for many scientists in other disciplines, it is also of interest to forensic scientists to summarise the information in a posterior distribution by a single number (a point value). A natural and regularly used summary of the posterior distribution is the posterior mean, though there are other ways to summarise the posterior distribution (e.g. the posterior median or the posterior mode). On the other hand, once it is accepted that uncertainty about a parameter may be represented by a probability distribution, it is a straightforward matter to determine a probability interval for the parameter. In other words, to determine a range of values (the so-called *Bayesian credible interval* or *probability interval*) such that the probability that θ belongs to that region is equal to a given amount $1 - \alpha$ (the so-called *credible level* or *credible probability*)

with $0 < \alpha < 1$. The interval is not uniquely defined. Any interval such that the probability that the parameter lies between the lower and the upper bound is equal to the credible level $(1 - \alpha)$ will result in a so-called $100(1 - \alpha)\%$ Bayesian credible interval. The shortest credible interval for θ is known as the *highest posterior density* (HPD) interval, and it is a subset C of the parameter set Θ such that

$$\int_C f(\theta \mid x) d\theta = 1 - \alpha.$$

The term *highest posterior density* is used to reflect that the interval chosen for a particular credible level, say, 0.95, corresponding to $\alpha = 0.05$, is the one for which the posterior probability density function takes its highest values whilst ensuring that the total probability within the interval is 0.95. The upper and the lower tails of the distribution do not necessarily have equal tail areas, as posterior distributions are not necessarily symmetric. In some cases, the HPD region consists of disjoint intervals (e.g. when the parameter θ is a proportion and the mass of the posterior distribution is spread around values that are either close to 0 or 1). Bayesian credible intervals are sometimes called Bayesian confidence intervals, but must not be confused with frequentist confidence intervals. A 95% credible interval allows the scientist to say that their degree of belief that the parameter θ lies in the realised range of values is equal to 0.95. Within the frequentist

paradigm, such an assertion is not valid as there is no probability distribution associated with the parameter. It is in fact no accident that the word probability is not used to describe a confidence interval. A frequentist confidence interval derives its validity as a method of inference on a long-run frequency interpretation of probability. Randomness is associated not with the parameter θ but to the lower and the upper bounds of the interval (sometimes called 'random' interval before observations become available). The probability with which a particular 95% confidence interval contains the parameter θ is not known. Suppose the experiment that generated the 95% confidence interval is repeated many times under identical conditions (a theoretical stipulation which is impossible to fulfil in practice), and on each of these occasions a 95% confidence interval is calculated. Then it can be said 95% of these (95%) confidence intervals will contain the parameter θ. This does not provide information concerning the one 95% confidence interval, which has been calculated. It is not known whether it does or it does not contain the parameter, and it is not even possible to determine the probability with which it contains the true value. Discussions on confidence intervals are not new in the judicial context (see e.g. (Kaye, 1987b); (Kaye and Freedman, 2002)).

There is one last argument that needs to be addressed. Consider again the issue of point estimation. It may be observed that there is not

a specific typical number that can be taken as a representative of the value of the parameter of interest. Sometimes there may be more than one typical number, as posterior distributions are not necessarily unimodal. The question of interest may be formulated as follows:

> *Should the estimate chosen be the value with the highest probability (the mode), a value such that the odds are one-to-one of the true parameter being above or below it (the median), a middle value in a centre of gravity (the mean), or something else? (Antelman, 1997, p. 356)*

This issue can be resolved within the decision-theoretic approach to the Bayesian statistical inference. The practising decision maker is asked to consider the expected loss of making a decision about the unknown parameter. The optimal decision is the one that minimises the loss incurred whenever a decision is taken. The core elements of the decision theoretic approach useful to address topics of statistical inference that are covered in this chapter are introduced in Section 4.7, as well as some standard loss functions that might be used in practice (Section 4.7.1). An application to the problem of sampling is given in Section 4.7.2.

In this chapter a detailed discussion of the Bayesian inference and decision analysis in the context of statistical models for univariate random variables is presented. An extension of the Bayesian inference to the multivariate case will be addressed in Chapter 7.

4.2 INFERENCE FOR A PROPORTION

A common need in forensic practice is to investigate the proportion of individuals or items that share a given target characteristic; for example, the target characteristic may be the true proportion of illicit drugs in a consignment. In general statistical terminology, the proportion is the parameter and, for general discussion, it is denoted θ. In such settings, experiments that reveal the presence or absence of the target characteristic can be seen as a sequence of Bernoulli trials (Section A.2.2), where each trial gives rise to one of two possible outcomes, conventionally labelled as success and failure. A statistical model for data that arise from a sequence of Bernoulli trials is represented by the binomial distribution $Bin(n, \theta)$ (see Section A.2.3), so that

$$\Pr(X = x \mid n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}; \quad x = 0, 1, \ldots, n, \tag{4.2}$$

where X is the random variable that represents the number of successes in each trial, n is the number of trials and θ the probability of success in each trial, or equivalently the proportion of successes in the population, and $\Pr(X = x \mid n, \theta)$ represents the probability of x successes. where n, the number of trials, and θ are both parameters.

However, it is only θ about which inference is desired. The number n of trials is fixed in advance of the data collection. Equation (4.2) may also be viewed as a likelihood function for θ and denoted $L(\theta \mid n, x)$. Within a Bayesian paradigm, θ has a probability distribution and it is desired to determine this in order to make inferences about θ. The most common prior distribution for θ is the beta distribution, denoted $Be(\alpha, \beta)$, a continuous distribution parameterised by α and β, with probability density function given by

$$f(\theta \mid \alpha, \beta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha, \beta); \quad 0 < \theta < 1 \tag{4.3}$$

(Section A.3.7). The beta distribution can take a variety of shapes (see Figure A.5). Hence, it can portray a variety of prior opinions reasonably accurately. The posterior distribution of θ can be obtained by an application of Bayes' theorem (4.1) and is given by

$$f(\theta \mid \alpha, \beta, x, n) = \frac{\theta^{\alpha+x-1}(1-\theta)^{(\beta+n-x-1)}}{B(\alpha+x, \beta+n-x)}; \quad 0 < \theta < 1, \tag{4.4}$$

denoted $Be(\alpha + x, \beta + n - x)$. In the particular case where $x = n$, the density function is given by

$$f(\theta \mid \alpha, \beta, n) = \theta^{\alpha+n-1}(1-\theta)^{\beta-1}/B(\alpha+n, \beta), \quad 0 < \theta < 1. \tag{4.5}$$

Similarly, when $x = 0$,

$$f(\theta \mid \alpha, \beta, n) = \theta^{\alpha-1}(1-\theta)^{\beta+n-1}/B(\alpha, \beta+n), \quad 0 < \theta < 1. \tag{4.6}$$

Note from (4.4) that if α and β are small relative to n, x, and $n - x$, then the choice of the parameters for the prior distribution is not very important. This is exemplified in Section 4.3.1 for the choice of sample size when sampling from consignments of discrete units such as tablets, compact disks, or video tapes. An application to the detection of nandrolone metabolites in urine is described in Robinson et al. (2001).

The beta prior distribution and binomial distribution combine together to give a posterior distribution, which is also a beta distribution, where parameters α and β are simply updated according to certain updating rules. The posterior distribution is a beta distribution as in (4.4) with updated parameters α' and β', where $\alpha' = \alpha + x$ and $\beta' = \beta + n - x$. See, e.g. Lee (2012) for further details. Note that the posterior distribution $Be(\alpha', \beta')$ becomes a new prior distribution that can be updated as new data become available. The beta distribution is a so-called *conjugate prior* distribution for the binomial distribution. A conjugate family is a set of prior distributions such that, for a particular probability distribution (e.g. the binomial distribution) and for every prior distribution in the family (e.g. for any beta prior distribution), the posterior distribution is also in

the same family as the prior distribution. A list of the more common conjugate families can be found in Bernardo and Smith (2000).

In some inference problems, it may happen that the information available prior to data collection is very limited, or that it is desired to minimise the impact of an expert's prior beliefs. In such situations, a so-called *vague prior* on which many observers may agree can be chosen. For the case of a beta distribution, this can be achieved by choosing $\alpha = \beta = 1$. This is a *uniform* prior for which $f(\theta \mid 1, 1) = 1$ for $0 < \theta < 1$ from (4.3). This is taken to represent prior ignorance of the value of θ as the function takes a constant value, 1, over the range of possible values of θ. One implication of this is that any interval of values between 0 and 1 has the same probability as any other interval of the same width.

An application to blood group frequencies is given in Weir (1996), with reference to data from Gunel and Wearden (1995). The parameter, θ, of interest is the frequency of allele M in the MN blood group system. From previous samples, Gunel and Wearden (1995) have a prior $Be(61, 44)$ distribution ($\alpha = 61, \beta = 44$ in the notation of (4.4)) for θ. The frequency $x = 201$ of the M allele in a sample of size $n = 372$ is taken as the data for the likelihood, from Race et al. (1949). Assuming Hardy–Weinberg equilibrium, the posterior distribution is $Be(61 + 201, 44 + 171) = Be(262,215)$. The prior and posterior density functions and the likelihood function are

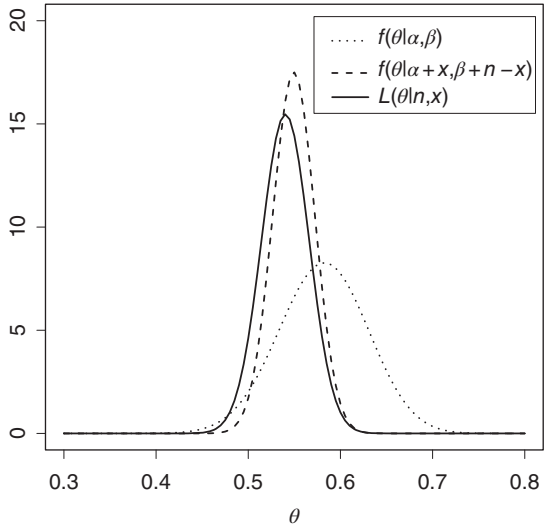

Figure 4.1 Prior density function $f(\theta \mid \alpha, \beta)$ with $\alpha = 61$, $\beta = 44$, $Be(61, 44)$ (dotted line), likelihood function $L(\theta \mid n, x)$ with $n = 372, x = 201$, $Be(201 + 1, 372 - 201 + 1)$ (solid line), and posterior density function $f(\theta \mid \alpha, \beta, x, n)$, $Be(61 + 201, 44 + 372 - 201)$ (dashed line), for a proportion. Source: Adapted from Weir (1996).

plotted in Figure 4.1. Note that the likelihood function is scaled (standardised), so that the area under the curve is 1. It can be shown that the scaled likelihood is a beta density with parameters $\alpha = x + 1$ and $\beta = n - x + 1$ (Taroni et al., 2010). The posterior density function $f(\theta \mid 262, 215)$ is more narrow, less dispersed (standard deviation is 0.023), than the prior density function $f(\theta \mid 61, 44)$ (standard deviation is 0.048), showing that the posterior density function has more

precise information about θ. This is the effect of the amount of information provided by the likelihood. It can be easily observed that as the amount of information increases, the posterior mean $(\alpha + x)/(\alpha + \beta + n)$ will depend more and more on the sample size and less and less on the prior mean, whilst the posterior variance will become smaller and smaller (see the mean and the variance of a beta distribution in Section A.3.7 with updated parameters $\alpha' = \alpha + x$ and $\beta' = \beta + n - x$).

There are instances where the observed outcomes of an experiment may well be more than two mutually exclusive outcomes. A statistical distribution for the number of observations that fall into each category is the multinomial distribution (Section A.2.4), whilst the conjugate prior distribution is the Dirichlet distribution (Section A.3.8). Consider the case of questioned printed documents, where control documents originating from a known source (a printer) are available for comparison. The polymer resins contained in dry black printer toner present on these documents can be analysed by means of Fourier transform infrared spectroscopy (FTIR), the results of which (so-called IR data) may be classified in one of several mutually exclusive categories. A Dirichlet-multinomial distribution (Section A.2.7) for the evaluation of black toner analyses in forensic document examination has been proposed by Biedermann et al. (2009b, 2011b).

4.2.1 Interval Estimation

Inference about θ, given the prior distribution $Be(\alpha, \beta)$ and the data (n, x) – under the assumption that sampling may be taken as with replacement, a reasonable assumption for large n – may be made with reference to the posterior beta distribution $Be(\alpha', \beta')$ in (4.4), where $\alpha' = \alpha + x$ and $\beta' = \beta + n - x$. Thus the shortest $100(1-\alpha)\%$ probability interval may be determined by determining the shortest interval in $(0, 1)$ within which $100(1-\alpha)\%$ of the distribution lies. Now, if the posterior density is roughly symmetric, the HPD interval will also be approximately symmetric and can be found simply by choosing equal tail areas as

$$[q_{\alpha/2}, q_{1-\alpha/2}]$$

where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of a Beta distribution $Be(\alpha, \beta)$ so that

$$\Pr\{q_{\alpha/2} \leq \theta \leq q_{1-\alpha/2}\} = 1 - \alpha.$$

These quantities are usually referred to as *quantiles* (Section A.3.1). Symmetry implies that intervals $(0, q_{\alpha/2})$ and $(q_{1-\alpha/2}, 1)$ are of equal length (see Figure 4.2a).

The required quantiles can be easily computed by common statistical software (such as `R`). Note that for parameter values sufficiently large, say, $\alpha > 10$ and $\beta > 10$, the beta distribution can be

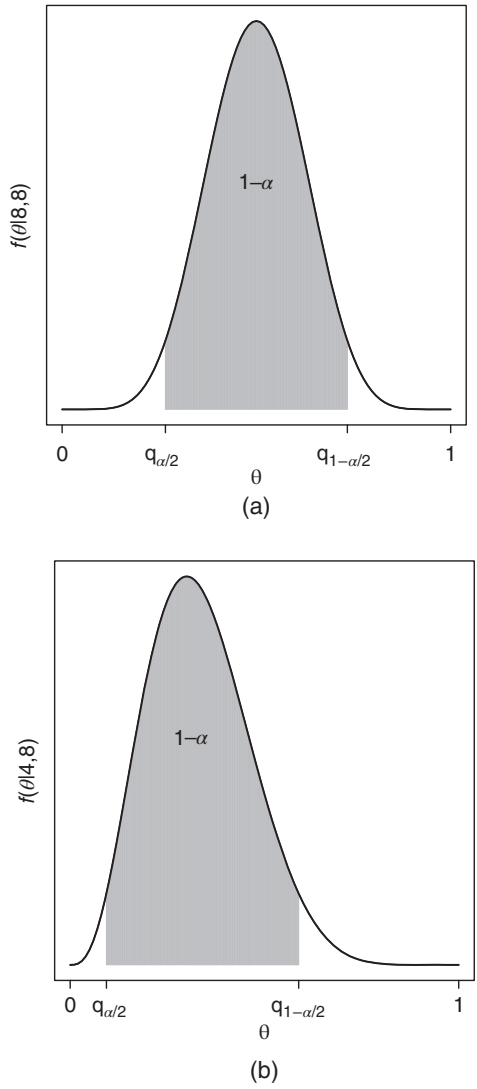

Figure 4.2 $100(1-\alpha)\%$ probability interval for a beta distributed random variable. (a) $Be(8,8)$; (b) $Be(4,8)$.

approximated by a Normal distribution (see, e.g. Bolstad and Curran (2017)) with mean

$$\mu = \frac{\alpha}{\alpha + \beta},$$

and variance

$$\nu = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

It follows that

$$\Pr\left\{ -z_{1-\alpha/2} \leq \frac{\theta - \mu}{\nu^{1/2}} \leq z_{1-\alpha/2} \right\} \approx 1 - \alpha,$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ point of a standard Normal distribution. The approximate $100(1-\alpha)\%$ credible interval for the proportion θ is

$$[\mu \pm z_{1-\alpha/2}\nu^{1/2}].$$

Consider a 95% credible interval. The 97.5% point of a Normal distribution is 1.96, thus one has approximately a 95% probability that the proportion lies within 1.96 (posterior) standard deviations of the (posterior) mean. Note that it is legitimate to refer to this interval as a probability interval.

However, in many occasions the beta posterior distribution is asymmetric, therefore a symmetric or equi-tailed interval will no longer be the best choice as it will be possible to find intervals having the same credible level but with smaller size. A $100(1-\alpha\%)$ HPD interval can be calculated

numerically (see, e.g. Taroni et al. (2010)). A simple way consists in starting from a small interval around the posterior mode (i.e. assuming unimodality, the point where the posterior density is high) and then progressively considering larger intervals (i.e. with decreasing posterior density) until the area underlying the posterior density between the lower and the upper limit of such an interval equals the desired credibility level $(1 - \alpha)$ (see Figure 4.2b). As $q_{\alpha/2}$ and $q_{1-\alpha/2}$ denote the $100\alpha/2\%$ and $100(1 - \alpha/2)\%$ quantiles and the distribution is asymmetric, the intervals $(0, q_{\alpha/2})$ and $(q_{1-\alpha/2}, 1)$ are not of equal length.

It is also possible to obtain a confidence interval for the true proportion in a binomial model. Denote the sample proportion by $\hat{\theta} = p = x/n$. The distribution of the sample proportion for large n can be approximated by a Normal distribution with mean equal to the population mean θ and variance equal to the population variance $\theta(1 - \theta)/n$ (Section A.3.2.1). It follows that

$$\Pr\left\{-z_{1-\alpha/2} \leq \frac{p - \theta}{(\theta(1 - \theta)/n)^{1/2}} \leq z_{1-\alpha/2}\right\} \approx 1 - \alpha,$$

and the approximate $100(1 - \alpha)\%$ confidence interval for θ is

$$[p \pm z_{1-\alpha/2}(p(1 - p)/n)^{1/2}]. \tag{4.7}$$

Note that in (4.7) the sample standard deviation $(\theta(1 - \theta)/n)^{1/2}$ is estimated by $(p(1 - p)/n)^{1/2}$.

Consider a 95% confidence level, so that the quantile is again 1.96. One can be 95% confident that the true proportion lies between 1.96 standard deviations of the sample mean. A credible (probability) interval and the corresponding confidence interval for a proportion θ might be approximately identical.

Consider drug sampling. A sample of n tablets from a large consignment is taken and x are found to be illicit. An interval may be determined for the true proportion of illicit tablets in the consignment. As an example, consider $n = 100$ and $x = 50$, and a uniform prior distribution with $\alpha = \beta = 1$. The posterior distribution being $Be(\alpha' = 51, \beta' = 51)$, the 95% HPD interval will be

$$[q_{0.025}, q_{0.975}] = [0.404, 0.596],$$

where $q_{0.025}$ and $q_{0.975}$ are the 2.5% and 97.5% points, respectively, of the beta posterior distribution that may be obtained from statistical software such as R. Conversely, the approximate 95% confidence interval will be

$$\begin{aligned}&[0.5 \pm 1.96 \times (0.5 \times 0.5/100)^2] \\ &\quad = [0.402, 0.598].\end{aligned}$$

However, the philosophies underlying their respective constructions are very different.

There is another approach to interval estimation based on the likelihood function. Consider

a large population of unrelated individuals in which it is desired to determine the proportion γ of people of blood group Γ. A sample of size n is taken. The sample is sufficiently small with respect to the size of the population that the proportion γ is effectively unchanged by the removal of the sample. The number X of people of blood group Γ in a sample of size n is a random variable with a binomial distribution, $X \sim Bin(n, \gamma)$ (Section A.2.3). Suppose a sample of size 30 is taken in which it is observed that 6 are of group Γ. The result in (4.7) can be reformulated as

$$[x/n \pm z_{1-\alpha/2}(x(n-x)/n^3)^{1/2}], \tag{4.8}$$

and the corresponding 95% confidence interval is

$$\begin{aligned}\frac{6}{30} \pm 1.96\sqrt{\frac{6(30-6)}{30^3}} &= 0.2 \pm 1.96\sqrt{\frac{0.2 \times 0.8}{30}} \\ &= 0.2 \pm 0.143 \\ &= [0.057, 0.343].\end{aligned}$$

This may be subject to the usual criticisms levelled against confidence intervals.

Another approach known as the likelihood approach considers as an interval estimate of γ, known as a *likelihood interval* all values of γ for which the likelihood function $L(\gamma \mid n, x)$ is greater than some fraction of $L(\hat{\gamma} \mid n, x)$, where $\hat{\gamma}$ is known as the *maximum likelihood estimate* (that is, the value γ at which the likelihood function attains its maximum). Examples quoted by Royall (1997)

for the fraction are $1/8$ and $1/32$; the interval estimate of γ corresponds to those values of γ for which $L(\gamma \mid n, x) > L(\hat{\gamma} \mid n, x)/8$ and for which $L(\gamma \mid n, x) > L(\hat{\gamma} \mid n, x)/32$, respectively. The values 8 and 32 are suggested by Royall (1997) as values to define the interval estimates of parameters such that for values of γ lying outside the intervals the data provide 'fairly strong' or 'strong' evidence in support of $\hat{\gamma}$ over γ. The use of such adjectives for the support for $\hat{\gamma}$ is not to be confused with similar adjectives given in Table 2.8 for likelihood ratio values. Consider the value 8, and values of γ for which $L(\gamma \mid n, x) > L(\hat{\gamma} \mid n, x)/8$. These form the likelihood interval (γ_1, γ_2) (see Figure 4.3). If γ lies in the interval (γ_1, γ_2) then there is no alternative value of γ for which the observations (x, n) (such that $\hat{\gamma} = x/n$) represent 'fairly strong' evidence in favour of $\hat{\gamma}$ rather than γ. For a value of γ outside (γ_1, γ_2) there is at least one alternative value, namely, $\hat{\gamma}$ that is better supported by a factor greater than 8. The end-points of the horizontal line in Figure 4.3 are (γ_1, γ_2) for the example $x = 6, n = 30$. For this example, the maximum value of $L(\gamma \mid n, x)$ is

$$L(0.2 \mid 30, 6) = \binom{30}{6} 0.2^6 0.8^{24} = 0.180.$$

The end-points of the interval for 'fairly strong' evidence are the values (γ_1, γ_2) for which $L(\gamma_1 \mid 30, 6) = L(\gamma_2 \mid 30, 6) = L(0.2 \mid 30, 6)/8 = 0.022$. These points may be verified to be $\gamma_1 = 0.081$ and $\gamma_2 = 0.372$. There is, thus, 'fairly strong' evidence from a sample of size 30 in which there are 6 'successes' that the true proportion

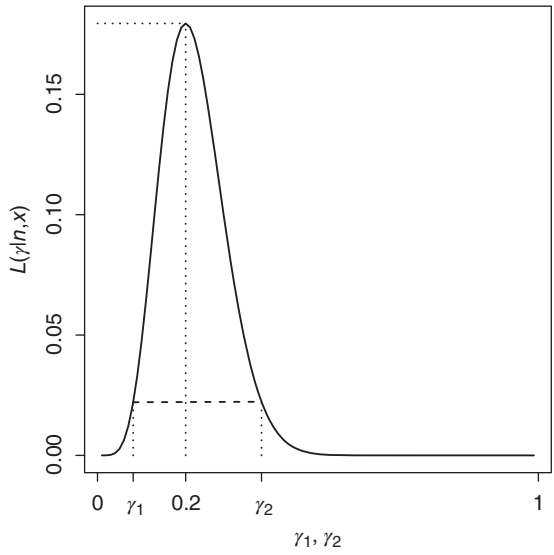

Figure 4.3 Likelihood function $L(\gamma \mid n, x)$ for the proportion of people of blood group γ, from a sample of size 30 in which 6 were of group γ. A likelihood interval, derived from the observed data, of fairly strong support for the values of γ included within it is indicated with the dashed line: $(\gamma_1 = 0.081, \gamma_2 = 0.372)$.

of successes in the population from which the binomial sample was drawn is in the interval $(0.081, 0.372)$. An interval determined by 32 (instead of 8) has a similar interpretation but with 'strong' replacing 'fairly strong'.

4.2.2 Estimation with Zero Occurrences in a Sample

In forensic science applications there are often situations where zero occurrences of a characteristic of interest are observed. For example, this

may be so in DNA-related matters, but it also may happen in relation to other types of scientific evidence. Consider the following scenario.

> *The scientist found a match in several physical properties between glass from a broken crime scene window and glass fragments found in connection with a suspect. He then surveyed 64 glass objects and found that none of these other 64 samples agreed with the glass found in his case. (Stoney, 1992, p. 383)*

There is no match with the case sample in a sample of 64 objects. Elicitation of the probability of a match by the sample relative frequency alone would give a value of zero and a likelihood ratio of infinity. However, the sample is only of size 64. Intuitively, this is insufficient to say that the glass in the crime scene window was unique.

An upper bound on the probability for the proportion of positive outcomes (e.g. characteristics of glass, illicit tablets, pirated CDs) in a population when there has been no observation of a positive outcome in a sample of n members of the population may be obtained using a Bayesian model with a beta prior and a binomial likelihood. Let θ denote the true, but unknown, proportion of positive outcomes in the population from which the sample has been taken. The prior distribution for θ is taken to be a beta distribution with parameters $\alpha = \beta = 1$ corresponding to a uniform prior. The likelihood is a binomial likelihood with n trials and the probability of a positive outcome is θ. In this example, there are no positive outcomes.

The posterior probability distribution for θ is then a beta distribution with parameters $x = 0$, n and $\alpha = \beta = 1$ in (4.6). The probability density function is

$$f(\theta \mid 1, 1, n) = \theta^0(1-\theta)^n / B(1, 1+n)$$
$$= (n+1)(1-\theta)^n.$$

The probability that $\theta > \theta_0$ is

$$\Pr(\theta > \theta_0 \mid 1, 1, n) = \int_{\theta_0}^{1} (n+1)(1-\theta)^n d\theta$$
$$= [-(1-\theta)^{n+1}]_{\theta_0}^{1} = (1-\theta_0)^{n+1}.$$

In the example from Stoney (1992), $n = 64$ and $\Pr(\theta > \theta_0) = (1-\theta_0)^{65}$.

A probabilistic upper bound θ_0 for θ may then be determined by choosing a probability level ϵ, say, and solving the equation

$$\epsilon = \Pr(\theta > \theta_0) = (1-\theta_0)^{n+1}. \qquad (4.9)$$

The solution can be easily obtained by a manipulation of (4.9)

$$\epsilon^{1/(n+1)} = (1-\theta_0),$$

so that

$$\theta_0 = 1 - \epsilon^{1/(n+1)}.$$

Note that the solution, not surprisingly, depends on the sample size n in the current survey. Let $n = 64$. When $\epsilon = 0.05, \theta_0 = 0.0450$ and when

$\epsilon = 0.01, \theta_0 = 0.0684$. Thus, in the context of the example from Stoney (1992), it can be said that

- there is a probability of 0.95 ($\epsilon = 0.05$) that the true proportion of matches in the population from which the crime scene window, the PoI's fragments and the 64 surveyed objects were taken is less than 0.0450;
- there is a probability of 0.99 ($\epsilon = 0.01$) that the true proportion of matches in the population from which the crime scene window, the PoI's fragments and the 64 surveyed objects were taken is less than 0.0684.

Solutions for an increasing sample size and two values of ϵ (0.01, 0.05) are displayed in Figure 4.4.

Note that in the special case of zero occurrences, the corresponding confidence interval is approximately equal to the Bayesian probability interval. The $100(1-\epsilon)\%$ confidence interval is of the form $[0, \theta_0]$ where θ_0 is the maximum θ such that $\Pr(X = 0 \mid n, \theta) > \epsilon$ so $\Pr(X = 0 \mid n, \theta) > \epsilon = (1-\theta)^{n+1}$. Thus the equation to solve is $(1-\theta)^{n+1} = \epsilon$, the same as (4.9).

Zero occurrences of an event are often observed with sequence analyses of human mitochondrial DNA (mtDNA); mtDNA is widely used to characterise forensic biological specimens, particularly when there is insufficient nuclear DNA in samples for typing.

The sequences of an mtDNA region from a recovered and control sample are compared. If the sequences are unequivocally different, then

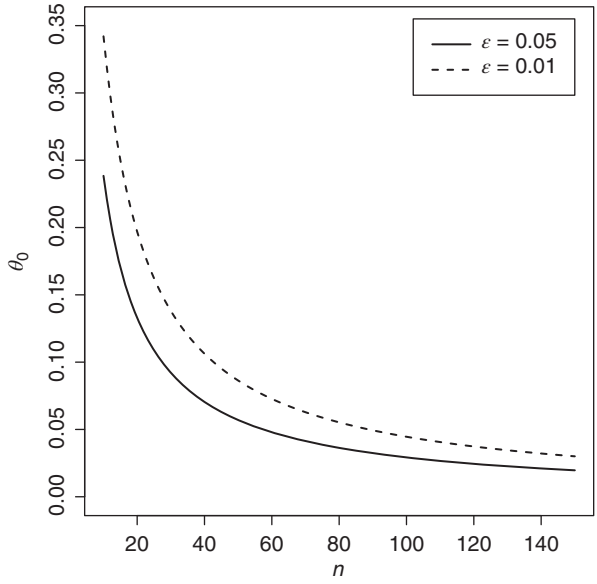

Figure 4.4 Probabilistic upper bounds θ_0 for the proportion θ of occurrences of an event in a population when there have been no occurrences in a sample of size n, for given values of $\epsilon = \Pr(\theta > \theta_0)$ and an increasing sample size n, with a beta prior for which $\alpha = \beta = 1$.

the sample can be excluded as originating from the same source. If the sequences are the same, then the recovered and control samples cannot be excluded as potentially being from the same source (note that no problems of nucleotide differences, mutations, or heteroplasmy are considered in what follows).

When there is no difference between the two samples, it is desirable to convey some information about the value of the evidence. Presently, the practice is to count the number of times a

particular sequence (or haplotype) is observed in a relevant database. Database allele frequency estimates are subject to several sources of uncertainty, both genetically and statistical. These include sampling variation and violation of the assumption of independence across loci. Ignoring these sources of uncertainty in the assessment of evidence might lead to an overestimation of the value of evidence. Several methods have been suggested for assessing sampling uncertainty, including a modification of the product rule in terms of a correction factor (Balding and Nichols, 1994; Balding and Steele, 2015) and so-called bootstrap methods (Curran et al., 2002). A review can be found in Buckleton et al. (2016c). A measure of sampling error is provided for these estimates, which is especially necessary when estimates are based on samples of just a few hundred profiles. This approach allows the scientist to communicate the value of the mtDNA evidence using the reciprocal of the relative frequency as a likelihood ratio (Carracedo et al., 2000).

Analogously, the use of Y-chromosome STR polymorphisms has become commonplace in forensic laboratories, e.g. Willuweit and Roewer (2007), Buckleton and Myers (2014), and Caliebe and Krawczak (2018). Their applications include the use in deficiency paternity testing cases (e.g. where the father is not available for analysis and inferences are made by reference to relatives) and, especially, the discrimination of stains in forensic investigations when a person of interest

is male (e.g. a male–female mixture in sexual assault cases). The counting method where the number of observations may be declared in a relevant database plays an important role in the assessment of the evidence (Gusmão et al., 2006; Scientific Working Group on DNA Analysis Methods (SWGDAM), 2009; Buckleton et al., 2011; Roewer and Geppert, 2012).

If there are no observations of a particular profile, in either the literature or the current survey, then a probabilistic upper bound on the profile may be obtained as described in this Section for the glass example of Stoney (1992).

Other relevant references for the estimation of a proportion include Louis (1981) for confidence intervals for a binomial parameter after observing no successes, Kaye (1987b) for a note on the burden of persuasion, Balding and Nichols (1994) and Curran et al. (2002) for examples in DNA profiling, and Taroni et al. (2010) for an example related to the percentage of banknotes contaminated with an illicit drug.

4.2.3 Uncertainty on Sensitivity and Specificity

Consider the medical diagnosis scenario that was described in Section 2.2.2, where the sample proportion of patients with a disease whose blood test was positive had been used to estimate the sensitivity of the test, and the sample proportion

of patients without such a disease whose blood test was negative had been used to estimate the specificity of the test. Sensitivity and specificity provide a measure of the quality of a test, with high values implying high quality. These values could be used, e.g. to make decisions or to answer specific questions about the condition of a specific patient. For example, given values in Table 2.3, it was possible to quantify the probability that a person is affected by a given disease (S) given a positive blood test (R).

However, such values are not observable in practice, and the true values related to the population of all patients are effectively unknown. The connection between the sample (observable) and the population (unknown) proportion can be made with reference to the beta-binomial statistical model described in Section A.2.7.

First, consider sensitivity, $\Pr(R \mid S)$. As it is reasonable to assume that the outcome of a test on a given patient with a given disease is not informative about the likely outcome of such a test on another patient having the same disease, such an experiment can be modelled as a sequence of Bernoulli trials, where each trial (medical test) has only two possible outcomes: *positive* (with probability θ), and *negative* (with probability $1 - \theta$). The number n of patients with a given disease S whose test is positive can therefore be modelled by a binomial distribution, with parameter n and θ (the probability that the test is positive given that

the patient has the given disease). Parameter θ (i.e. the sensitivity of the test) is clearly unknown, and a beta prior distribution can be used to model prior uncertainty. There are two possible approaches for representing uncertainty about θ. A uniform prior distribution can be assumed, where all possible values of θ between 0 and 1 are *a priori* equally likely. This is equivalent to a $Be(1, 1)$ prior probability density. As an alternative, an informative prior distribution incorporating knowledge from sources other than the sample proportion can be incorporated. Consider for illustrative purposes, a uniform prior distribution $Be(1, 1)$, and values in Table 2.3. The posterior distribution of the sensitivity is still a beta distribution, with parameters updated as in (4.4), where $\alpha = 1 + 95$ (where 95 is the number of patients, out of 100, who have the disease and give a positive blood test) and $\beta = 1 + 5$ (where 5 is the number of patients, out of 100, who have the disease and give a negative blood test). The posterior distribution (Figure 4.5) has a mass concentrated on the value of the sensitivity obtained empirically using the sample proportion, with a spread reflecting the uncertainty about such a value. A credible interval may be reported as in Section 4.2.1. For example, there is a probability equal to 0.95 that the sensitivity takes values in the interval (0.89, 0.98) the lower and upper values that represent the quantiles of order 0.025 and 0.975, respectively, of a $Be(96, 6)$ distribution.

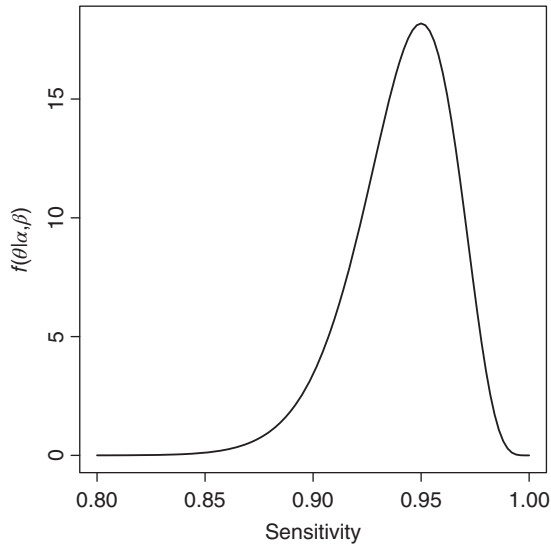

Figure 4.5 Posterior distribution $f(\theta \mid \alpha, \beta)$ of the sensitivity $Be(\alpha = 96, \beta = 6)$.

Second, the uncertainty about the specificity of the test can be modelled analogously. Given the population of patients not affected by such a disease, it is possible to model such an experiment as a Bernoulli trial, where each trial has only two possible outcomes: *negative* (with probability ϕ) and *positive* (with probability $1 - \phi$). Considering again a uniform prior distribution, the posterior distribution can be obtained as above with $\alpha = 1 + 99$ (where 99 is the number of patients out of 100 who do not have the disease and give a negative blood test) and $\beta = 1 + 1$ (where 1 is the number of patients out of 100 who do

not have the disease and give a positive blood test) (values for the number of successes and for the number of failures are again taken from Table 2.3).

The question of how a positive test can be used to update the probability that a patient testing positive is actually diseased was addressed in Example 2.1 of Section 2.2.2. Sensitivity and specificity were assumed to be known. Now, there is uncertainty about such values, as sample values in Table 2.3 are used to infer the unknown corresponding values in the entire population. Uncertainty about the effective value of the sensitivity and of the specificity of the test are represented by a probability distribution, and this implies that the posterior probability in (2.5) has itself a distribution.

One way to obtain such a distribution is to sample several pairs of values of the sensitivity and specificity of the test from the respective posterior distributions (in the specific case, from $Be(96, 6)$ and $Be(100, 2)$), and calculate for each pair the probability that the person is affected by the disease given that the medical test is positive (Parmigiani, 2002). In this way, a sample of values can be obtained and are represented in Figure 4.6. From Figure 4.6 it can be observed that the required probability could take a range of values from 0.5 to 1, though values in the right tail are more likely.

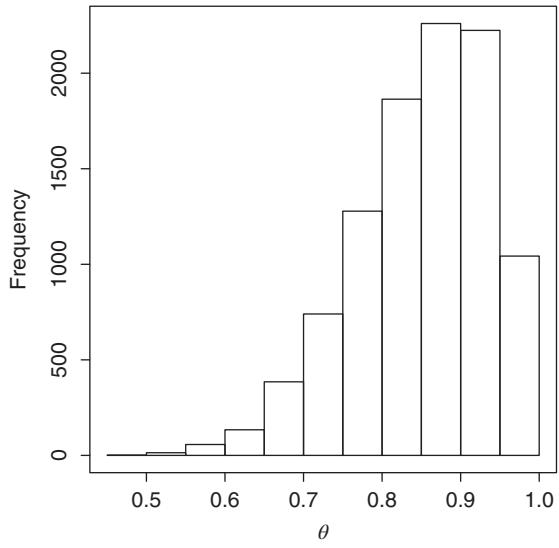

Figure 4.6 Histogram of the frequency of 1000 sampled values of the probability θ that a person with a positive medical test is affected by a given disease.

4.3 SAMPLING

Sampling issues represent a topic of ongoing interest to the forensic community essentially because of their crucial role in laboratory planning and working protocols. Forensic laboratories are commonly asked to inspect consignments of discrete units whose characteristics may be of interest within a criminal investigation. Typical examples include consignments such as bags or electronic storage devices, that consist of individual items such as pills or images. Each item in such a consignment may, or may not, contain something illegal (e.g. drugs, pornographic

images). It is of interest to an investigating scientist to determine the proportion of the consignment that contains something illegal. This may be done exactly (assuming no mistakes are made) by examination of every unit in the consignment. Such an examination can be extremely costly. Considerable resources can be saved if information, sufficient to satisfy the needs of investigators, may be gained from examination of a sample from the consignment. Uncertainty is introduced when inference is made from the sample to the population, because the whole population is not inspected. However, this uncertainty may be quantified probabilistically.

A general introduction to sampling techniques is given in Cochran (1977). Ideas on sample size determination are discussed in Smeeton and Adcock (1997). A very good discussion of various types of samples, such as random samples, representative samples, and convenience samples, is given in Evett and Weir (1998). A review of statistical and legal aspects of the forensic study of illicit drugs is given by Izenman (2001) This includes a discussion of various sampling procedures, various methods of choosing the sample size, a strategy for the assessment of homogeneity, and the relationship between quantity and the possible standards of proof. Further comments on sampling issues are given in Aitken et al. (1997), Biedermann et al. (2008b), Bring and Aitken (1997), Curran et al. (1998a), in various chapters of Gastwirth (2000), such as Aitken (2000),

Gastwirth et al. (2000), and Izenman (2000a,b,c) and in Izenman (2003). A relevant case is that of *U.S. v. Shonubi* (1992, 1995, 1997).

Only inferences from simple random samples are discussed here. It may be that it is not possible to take a simple random sample. If so, the following comments are still of relevance. The comments are made in the context of sampling for the estimation of allelic frequencies in DNA profiles but are applicable to other areas of forensic science, including drug sampling which is the main example in this section.

> *Of course, a real crime laboratory would not attempt . . . to take a random, representative, stratified sample of individuals to address the question of issue. In the vast majority of cases, the laboratory will have one or more* convenience *samples. Such a sample may be of laboratory staff members, or from blood donor samples with the cooperation of a local blood bank, or from samples from victims and suspects examined in the course of casework.*
>
> *. . . [In] the forensic context, we will generally be dealing, not with random but with convenience samples. Does this matter? The first response to that question is that every case must be treated according to the circumstances within which it has occurred, and the next response is that it is always a matter of judgement. . . . In the last analysis, the scientist must also convince a court of the reasonableness of his or her inference within the circumstances as they are presented as evidence. (Evett and Weir, 1998, pp. 44–45)*

Several sampling procedures, including random sampling, are discussed in Izenman (2001). First, for single containers, examination by a chemist of a random sample of a substance seized within

a single bag or container has been accepted by the courts to prove the identity of the remainder of the substance in the container. For multiple containers, without homogeneity, a rule is that at least one sample from each container should be conclusively tested for the presence of an illicit drug. Another procedure is that of composite sampling. In this procedure, a sample is taken from each source, the samples are then thoroughly mixed and a subsample is taken from the mixture. The mixture is the composite sample.

This section concerns the choice of sample size and the interpretation of data from samples. The following questions in particular are addressed and answered:

- How big a sample should be taken?
- What proportion of a consignment of discrete, homogeneous items is illicit?

Both questions have probabilistic answers and these can be determined using a Bayesian framework. With reasonable assumptions, a probability distribution for the proportion of units in the consignment may be derived, based on the scientist's prior belief (i.e. prior to the inspection of individual units) and the outcome of the inspection of the sample. The strength of the scientist's prior beliefs may be expressed by a probability density function as described previously in Section 4.2. It is possible to choose the function in such a way that the effect of the scientist's prior beliefs is very small (or very large). The choice of the binomial

model which is used to represent the uncertainty introduced by the sampling process is a subjective choice influenced by the scientist's prior beliefs. The choice of the binomial distribution used here requires assumptions about independence of the probability for each unit being illegal and the choice of a constant value for this probability.

A comparison will be made of the results obtained from the Bayesian and frequentist approaches to the assessment of uncertainty, in order to (i) contrast the clarity of the inferences obtainable from the Bayesian approach with the lack of clarity associated with the frequentist approach and (ii) illustrate the greater flexibility of the Bayesian approach with the inflexibility of the frequentist approach.

The methods are illustrated with reference to sampling from consignments of drugs. However, they apply equally well to sampling in other forensic contexts, for example, glass fragments (Curran et al., 1998a) and pornographic images.

Frequentist procedures are described in Tzidony and Ravreboy (1992) for choosing a sample size from a consignment. A distinction is drawn between an approach based on the binomial distribution and an approach based on the hypergeometric distribution (Section A.2.5). It is argued in Tzidony and Ravreboy (1992) that the former can be used for large consignments in which the sampling of units may be considered equivalent to sampling with replacement. For small samples, the sampling units cannot be so

considered, sampling is without replacement and the hypergeometric approach is used (Frank et al., 1991). The Bayesian approach also has different methods for analysing large and small samples.

Various methods for selecting the size of a random sample from a consignment have been accepted by the US courts (Frank et al., 1991; Izenman, 2001). A summary of different procedures used in 27 laboratories around the world is given in Colón et al. (1993). Various procedures suggested for the choice of sample size include methods based on the square root of the consignment size, a percentage of the consignment size, and a fixed number of units regardless of the consignment size, as well as the hypergeometric distribution. The formula

$$m = 20 + 10\%(N - 20) \quad (\text{for } N > 20), \quad (4.10)$$

where m is the sample size, the number of items inspected, and N is the consignment size is proposed by Colón et al. (1993) and is listed by the UN Office on Drugs and Crime (2009) along with other so-called *arbitrary sampling* rules. As well as being simple to implement, this approach, as the authors rightly claim, provides the opportunity to discover heterogeneous populations before the analysis is completed. It has also been suggested that 'an inference made at the 95% confidence level, that 90% or more of the packages in an exhibit contain the controlled substance should be accepted as sufficient proof in such cases', Frank et al. (1991). These summaries are given

as confidence limits using a frequentist approach and not in probabilistic terms.

A Bayesian approach can provide summaries in probabilistic terms such as

> *'How big a sample should be taken for it to be said that there is a 100p% probability that the proportion on units in the consignment which contain drugs is greater than $100\theta_0$%?'*

or, for a particular case, with $p = 0.95$ and $\theta_0 = 0.50$,

> *'How big a sample should be taken for it to be said that there is a 95% probability that the proportion on units in the consignment which contain drugs is greater than 50%?'*

Some of the most commonly used approaches for sampling are compiled in a booklet issued by the ENFSI Drug Working Group (ENFSI, 2016).

Results similar to those discussed here for sample size determination and which take into account the limited size of the consignment and the discrete nature of the random variables are given by Weusten (2011) and Moroni et al. (2012). The methods described here assume the consignment is homogeneous. A method based on multiple sampling to take account of heterogeneity is described by Dujourdy et al. (2013).

4.3.1 Choice of Sample Size in Large Consignments

A large consignment is taken to be one which is sufficiently large that sampling is effectively with

replacement (Section A.2.3). This size can be as small as 50, though in many cases it will be of the order of many thousands.

A consignment of drugs containing N units will be considered a random sample from some super-population (Section 2.5.5) of units containing drugs. Let θ $(0 < \theta < 1)$ be the proportion of units in the super-population, which contain drugs. For consignment sizes of the order of several thousand all realistic values of θ will represent an exact number of units. For small sample sizes less than 50, θ can be considered as a nuisance parameter (i.e. one that is not of primary interest) and integrated out of the calculation leaving a probability distribution for the unknown number of units in the consignment which contain drugs as a function of known values. For intermediate calculations, θ can be treated as a continuous value in the interval $(0 < \theta < 1)$, without any detriment to the inference. Let m be the number of units sampled. The ratio m/N is known as the *sampling fraction*. Denote the number which are found to contain drugs by z.

A criterion has to be specified in order that the sample size may be determined. Consider the Bayesian criterion that the scientist wishes to be $100p\%$ certain that $100\theta_0\%$ or more of the consignment contains drugs when all units sampled contain drugs $(z = m)$. The criterion may be written mathematically as

$$\Pr(\theta > \theta_0 \mid \alpha, \beta, m, m) = p, \qquad (4.11)$$

or

$$\int_{\theta_0}^{1} \frac{\theta^{\alpha+m-1}(1-\theta)^{(\beta-1)}}{B(\alpha+m,\beta)} d\theta = p, \qquad (4.12)$$

using a beta conjugate prior distribution and a binomial distribution to give a beta posterior distribution (4.4), with the special case where the number of 'successes' equals the number of trials (4.5). The term $B(\alpha+m,\beta)$ is a beta function (A.22). Such integrals are easy to evaluate using standard statistical packages, such as R, given values for α, β, and m. Table 4.1 contains, for different values of α and β, and different values of m, the corresponding probabilities p satisfying (4.12) for $\theta_0 = 0.5$. Note from Table 4.1 that, for values of $m = 4, 5$, differing values of α and β, as long as both are small, have little effect on p.

Table 4.1 Probability that the proportion of drugs in a large consignment is greater than 50% for various sample sizes m and prior parameters α and β, $\Pr(\theta > 0.5 \mid \alpha, \beta, z, m)$.

α	β	m			
		2	3	4	5
1	1	0.88	0.94	0.97	0.98
0.5	0.5	0.92	0.97	0.985	0.993
0.065	0.935	0.78	0.90	0.95	0.97

Note that $z = m$.
Source: From Aitken (1999). Reprinted with permissions of ASTM International.

An alternative way of looking at (4.12) is to reverse the role of the parameters and solve for m. Given specified values for θ and p and values for α and β chosen according to prior knowledge, the appropriate value of m to solve (4.12) may be found, again using standard statistical packages. Consider the case where prior parameters α and β are set equal to 1. The sample size m (given that all items are found to be positive) required to be $100p\%$ certain that the proportion θ of positive units is larger than a specified threshold, say, θ_0, is then given by the value of m which satisfies the equation

$$\begin{aligned}\Pr(\theta > \theta_0 \mid 1, 1, m, m) &= \int_{\theta_0}^{1} \frac{\theta^m}{B(1+m,1)} d\theta \\ &= 1 - \theta_0^{m+1} = p, \quad (4.13)\end{aligned}$$

as $B(1+m,1) = 1/(m+1)$ (Section A.3.7). The value of m is thus given by the smallest integer greater than

$$[\log(1-p)/\log(\theta_0)] - 1. \quad (4.14)$$

This expression provides an answer to the question 'how many units should be inspected to satisfy this criterion?'. The dependency of the sample size on the values of p and θ_0 is illustrated in Table 4.2 for $p = (0.90, 0.95, 0.99)$ and $\theta_0 = (0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99)$. Take $p = 0.95$ and $\theta_0 = 0.5$. For large consignments, of whatever size, the scientist needs only examine four units,

Table 4.2 The sample size required to be $100p\%$ certain that the proportion of units in the consignment which contain drugs are greater than θ, when all the units inspected are found to contain drugs.

θ_0	p		
	0.90	0.95	0.99
0.5	3	4	6
0.6	4	5	9
0.7	6	8	12
0.8	10	13	20
0.9	21	28	43
0.95	44	58	89
0.99	229	298	458

The prior parameters $\alpha = \beta = 1$.
Source: From Aitken (1999). Reprinted with permissions of ASTM International.

in the first instance. This sample size is not large. However, there is not very much information gained about the exact value of θ. It has only been determined that, if all are found to contain drugs, there is a 95% probability that at least 50% of the consignment contains drugs.

Compare the clarity and flexibility of this result with that derived from a frequentist perspective. Consider the sample proportion $p = z/m$ which is an unbiased estimator of θ, i.e. $\mathrm{E}(Z/m) = \theta$ (see Section A.1). The variance of the sample proportion p is given by Cochran (1977) as

$$\frac{\theta(1-\theta)}{m}\left(\frac{N-m}{N-1}\right), \tag{4.15}$$

where $(N-m)/(N-1)$ is the *finite population correction factor* (fpc). The variance of the sample proportion that has been introduced in Section 4.2.1 for calculating a confidence interval for the proportion was based on the premise that observations are taken with replacement. However, in populations of finite size N, sampling cannot be equated to sampling with replacement, and the correction factor allows more precise estimates of θ to be obtained. Interestingly, provided that the sample size m is small in comparison with the population size N (say, less than 5% of the population is sampled), the size of the population has no direct effect on the precision of the estimate of θ. For example, if θ is the same in the two populations, a sample of 500 from a population of 200 000, gives almost as precise an estimate of the population proportion as a sample of the same size 500 from a population of 10 000. It can easily be shown using the result in (4.15) that the estimated standard deviation of θ in the second case is 0.98 times the estimated standard deviation in the first case. Little is to be gained by increasing the sample size in proportion to the population size. An increase of sample size by a factor of 20, from 10 000 to 200 000 gives a more precise estimate of the parameter of interest. However, the precision of the estimate in the sample of size 10 000 is 98% of that of the sample of 200 000. A large increase in sampling costs leads to only a very small increase in precision.

To simplify matters, assume the sampling fraction is small so that the fpc can be ignored and the

standard deviation of the sample proportion p is

$$\sqrt{\frac{\theta(1-\theta)}{m}}.$$

Suppose further the sample size and proportion are such that the sample proportion p may assumed to be approximately Normally distributed.[1] Recall from Section 4.2.1 that the width of the confidence interval for the proportion θ depends also on θ, which is unknown and can be estimated only after sampling. It is necessary to have some guess about the value of θ before determining the sample size. The expression $\theta(1-\theta)$ takes its maximum value of 0.25 at $\theta = 1/2$ and its minimum value of 0 when $\theta = 0$ or 1. A conservative choice of sample size is to take $\theta = 1/2$. Assume that θ is thought to be about 0.75. It is stipulated that a sample size m is to be taken to estimate θ to within 25%, that is, in the interval $(0.5, 1.0)$, with approximately 95% confidence. The criterion for the sample size is that there should be a confidence of 0.95 that the proportion lies in interval 0.75 ± 0.25. This implies that two standard deviations equal 0.25, that is,

$$2\sqrt{\frac{\theta(1-\theta)}{m}} = 0.25,$$

[1]The Normal approximation is accurate even if the proportion θ is close to 0 or 1 providing m is sufficiently large. However, even for samples of modest size, the approximation is still accurate provided that θ is close to 0.5 and the distribution is roughly symmetric.

(see Section A.3.2). This gives the following expression for m

$$m = \frac{4\theta(1-\theta)}{0.25^2}. \tag{4.16}$$

When $\theta = 0.75$, $m = 12$.

Thus, from a frequentist perspective, a sample of size 12 in which all sampled items are illicit is sufficient to estimate θ to be greater than 0.5 with confidence 0.95. Contrast this with the result derived from the Bayesian approach, which gave a value of 4 for the sample size.

The Bayesian methodology can be extended to allow for units that do not contain drugs. Take, for example, the case where one of the original four units is found not to contain drugs, then the posterior distribution is $Be(4, 2)$. Suppose three more units are inspected and all are found to contain drugs. It can be shown that the probability that $\theta > 0.5$, given that six out seven units contain drugs is larger than 0.95:

$$\begin{aligned}\Pr(\theta > 0.5 \mid 4, 2, 3, 3) &= \int_{0.5}^{1} \frac{\theta^{4+3-1}(1-\theta)^{2-1}}{B(4+3, 2)} d\theta \\ &= 0.96.\end{aligned}$$

Here $\alpha = 4, \beta = 2$ from the posterior probability distribution determined from the initial sampling with three out of four units found to be illicit. This distribution is taken to be the prior for the second sample where $x = n = 3$ in the notation of (4.4).

A sequential approach to sampling is described in Moroni et al. (2012). Take an initial sample. If

all members of the sample contain drugs, deem all members of the consignment to contain drugs. If not all members of the sample contain drugs, take an additional sample from the consignment to find a lower limit on the number of members of the consignment that contain drugs.

Obviously, when considering the results in Table 4.2, the consignment size has to be taken into account in order that the sample size may be thought small with respect to the size of the consignment. Thus, for the last row in particular to be useful, the size of the consignment from which the sample is to be taken will have to be of the order of several thousands.

There may be situations in which different choices of α and β may be more appropriate, though there may be concerns that it may be very difficult for a scientist to formalise their prior beliefs. An informal approach is described by Zamengo et al. (2011).

- If there is no prior belief about the contents of seizures from a consignment, set $\alpha = \beta = 1$, the uniform distribution.
- If there is a prior belief that either all items from the consignment contain drugs or no items at all contain drugs set $\alpha = \beta = 0.5$.
- If there is prior information that it is illicit drugs that are being considered and that all items contain drugs, the higher α should be relative to β.

A more formal approach to the determination of α and β requires the expert to assess a minimum

of two summaries. It may be the scientist has some substantial prior beliefs about the proportion of the consignment that may contain drugs to which they are prepared to testify in court. These beliefs may arise from previous experiences of similar consignments, for example. In such cases, use can be made of various properties of the beta distribution to assist the scientist in choosing values for α and β. When there is a substantial amount of information, summaries of location (e.g. the mean or expectation) and dispersion (e.g. the variance) that characterise the distribution of interest may be elicited from the expert. Denote by l the available location summary, and by d the available dispersion summary, respectively. Values for α and β can be chosen by matching these summaries to the corresponding moments (i.e. the mean and the variance) of the beta distribution (Section A.3.7), that is

$$l = \frac{\alpha}{\alpha + \beta},$$

$$d = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Solving these two equations for α and β gives estimates of the parameters of the $Be(\alpha, \beta)$ prior, that is

$$\alpha = \{l^2(1 - l)/d\} - l, \tag{4.17}$$

$$\beta = \{l(1 - l)^2/d\} - 1. \tag{4.18}$$

However, the information about the dispersion (*a priori*) may be unavailable. A practitioner (expert) may be able to provide an estimate of

the proportion (e.g. the sample proportion from a previous experiment), which can be used to assess the mean $\alpha/(\alpha+\beta)$ of the prior distribution, but at least one further quantity will be needed to elicit the full prior distribution. A method known as the *equivalent sample size method* can be implemented according to which the expert is requested to provide an estimate of the sample size upon which they are basing their assessment about the sample proportion. Recall the mean and the variance of the sample proportion are θ and $\theta(1-\theta)/n$. Parameters α and β of the beta distribution can be fixed by equating the mean and the variance of the sample proportion evaluated at p to the mean and the variance of the beta prior distribution, as

$$p = \frac{\alpha}{\alpha+\beta}, \tag{4.19}$$

$$\frac{p(1-p)}{n} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{4.20}$$

Solving the equations for α and β gives

$$\alpha = p(n-1), \tag{4.21}$$

$$\beta = (1-p)(n-1). \tag{4.22}$$

Consider as an example a case where a sample of glass objects is inspected and a given number of positive outcomes of some nature is observed. A beta prior distribution may be introduced to model uncertainty about the proportion of glass objects having this outcome. Prior knowledge may be given by a sample of $n = 40$ glass objects

collected in a previous experiment amongst which one positive result has been observed. Parameters α and β can be determined by substituting $n = 40$ and $p = 1/40$ in (4.21) and (4.22), to obtain, approximately, $\alpha = 1$ and $\beta = 38$. (The exact values are 0.975 and 38.025, respectively.) The prior density $Be(1, 38)$ is depicted in Figure 4.7. The prior mass is concentrated at low values of θ.

Alternatively, if it was felt that β could be set equal to 1 so that, for $\alpha > 1$, the beta function is monotonic increasing with respect to θ, and that there was a prior belief about a lower bound l for the proportion, say, that

$$\Pr(\theta > l \mid \alpha, \beta = 1) = p,$$

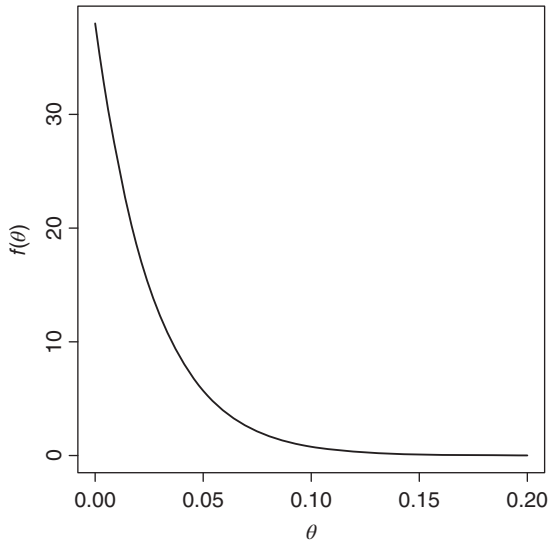

Figure 4.7 Beta prior distribution $Be(1, 38)$ over θ for $\theta \in (0, 0.20)$.

with l and p given, then

$$\alpha = \log(1-p)/\log(l).$$

Other elicitation methods have been developed for quantifying opinion about a proportion θ when the underlying model is a binomial distribution. A review can be found in O'Hagan et al. (2006) and references therein.

Practitioners should clearly inspect the shape of the elicited prior distribution to verify its consistency with their prior beliefs. Noting that with $\alpha/(\alpha+\beta) = p$ and $\beta/(\alpha+\beta) = (1-p)$, (4.20) can be rewritten as

$$\frac{p(1-p)}{n} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{p(1-p)}{\alpha+\beta+1},$$

so that $\alpha+\beta+1 = n$, where n represents the equivalent sample size. Bolstad and Curran (2017) warn that given the elicited prior distribution (e.g. as in (4.17) and (4.18)) one should compute the equivalent sample size to check whether the available knowledge is realistically comparable with the knowledge that would have been obtained from an experiment of that size. If not, they suggest one should increase the standard deviation and calculate a new prior, in order to avoid the assignation of a prior distribution that reflects more information than that which is effectively available.

Variation in the prior beliefs, expressed through variation in the values of α and β may have little

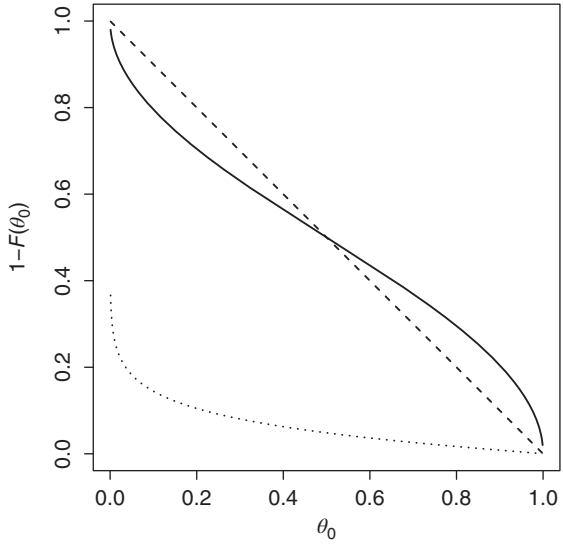

Figure 4.8 The prior probability $1 - F(\theta_0)$ that the proportion θ of units in a consignment is greater than θ_0, for various choices of α and β: $\alpha = \beta = 1$ (dashed curve), $\alpha = \beta = 0.5$ (solid), $\alpha = 0.065$, $\beta = 0.935$ (dotted). Source: From Aitken (1999). Reprinted with permissions of ASTM International.

influence on the conclusions, once some data have been observed. Figure 4.8 illustrates the prior probability that the true proportion of illegal units in a consignment is greater than a value θ_0, $0 < \theta_0 < 1$, for three choices of α and β producing a radically different shape of the prior distribution (see Figure A.5 in Section A.3.7), and therefore divergent prior beliefs. The values $\alpha = \beta = 1$ are chosen so that $\Pr(\theta > 0.5) = 0.5$ with a uniform distribution. The values $\alpha = \beta = 0.5$ also have $\Pr(\theta > 0.5) = 0.5$ but with more belief

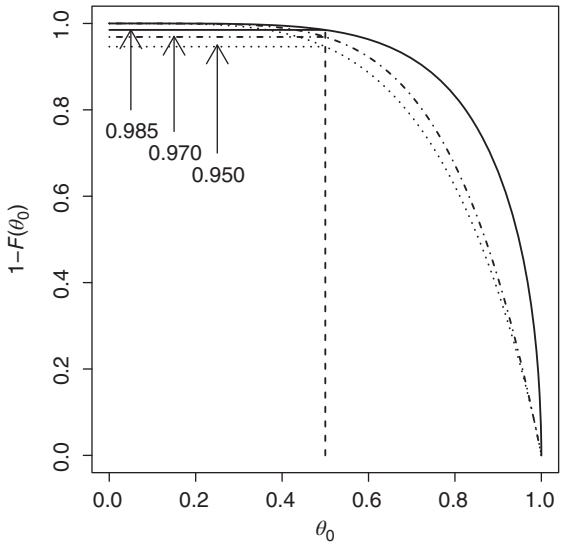

Figure 4.9 The posterior probability $1 - F(\theta_0)$ that the proportion θ of units in a consignment is greater than θ_0, for various choices of α and β: $\alpha = \beta = 1$ (dashed curve), $\alpha = \beta = 0.5$ (solid), $\alpha = 0.065, \beta = 0.935$ (dotted), after observation of four units all found to be illegal. The corresponding probabilities that at least 50% of the consignment contains illegal units is marked as 0.985 ($\alpha = \beta = 0.5$), 0.970 ($\alpha = \beta = 1$), 0.950 ($\alpha = 0.065, \beta = 0.935$). Source: From Aitken (1999). Reprinted with permissions of ASTM International.

in the tails of the distribution. The values $\alpha = 0.065, \beta = 0.935$ are chosen so that $\alpha + \beta = 1$ and $\Pr(\theta > 0.5) = 0.05$, the complement of the desired posterior probability that $\Pr(\theta > 0.5) = 0.95$. Figure 4.9 illustrates the posterior distribution that the true proportion of illegal units in a consignment is greater than θ_0, for these choices of α and β, once four units have been examined and all found to be illegal.

4.3.2 Choice of Sample Size in Small Consignments

Consider a consignment of N tablets that are homogeneous in nature (colour, texture, type of logo) and about which it is desired to learn more about the proportion which are illicit. Let R denote the number (unknown), out of N, which are illicit. A random sample of size m is examined and $z(\leq m)$ are found to be illicit. The probability of this event (z tablets found to be illicit when m tablets are chosen at random from N tablets of which R are illicit) is given by the hypergeometric distribution (Section A.2.5). A Bayesian approach for small consignments, using the hypergeometric distribution, is described in Aitken (1999), Coulson et al. (2001b), Weusten (2011), and Moroni et al. (2012). A discrete prior distribution is chosen for the $(N + 1)$ possible divisions of the consignment into licit and illicit terms. The likelihood function is based on the hypergeometric distribution sampling m from N and a posterior distribution obtained. A beta-binomial distribution (Section A.2.7) is used to provide a probability statement about the number of units in the consignment that contain drugs.

As before, let θ, satisfying $(0 < \theta < 1)$, be the proportion of illicit units in the super-population. The probability distribution of z, given m and θ, may be taken to be binomial, assuming a constant probability θ, independent of the sample size m. For each unit, independently of the others in the consignment, the probability it is illicit is taken to be equal to θ. The posterior distribution of θ is

another beta distribution, with parameters $(\alpha + z)$ and $(\beta + m - z)$.

Since the consignment size is small, a better representation of the variability of the number of illicit units in the non-inspected consignment is obtained considering a probability distribution for this number, Y, say, explicitly. Let there be n units in the reminder of the consignment (such that $m + n = N$), which have not been inspected. Then Y (unknown and no greater than n) is the number of illicit units in this remainder. Given θ, the distribution of $(Y \mid n, \theta)$, like that of $(Z \mid m, \theta)$, is binomial. However, θ has a beta distribution. Therefore, the distribution of $(Y \mid n, \theta)$ and the distribution of $(\theta \mid \alpha, \beta, m, z)$ can be combined to give a Bayesian predictive distribution for $(Y \mid m, n, y, \alpha, \beta)$, also known as a beta-binomial distribution (Section A.2.7):

$$\Pr(Y = y \mid m, n, z, \alpha, \beta) = \frac{\Gamma(m + \alpha + \beta)\binom{n}{y}\Gamma(y + z + \alpha)\Gamma(m + n - z - y + \beta)}{\Gamma(z + \alpha)\Gamma(m - z + \beta)\Gamma(m + n + \alpha + \beta)}, \quad (y = 0, 1, \ldots, n). \tag{4.23}$$

From this distribution, inference can be made about Y, such as probability intervals or lower bounds for Y.

As with large consignments, values for α and β may be chosen subjectively to represent the scientist's prior beliefs before inspection about

the proportion of the units in the consignment (as a random sample from the super-population) which contain drugs. Consider the beta-binomial distribution in (4.23) with $\alpha = \beta = 1$. It can be shown that

$$\Pr(Y = y \mid m, n, z, 1, 1) = \frac{(m+1)\binom{m}{z}\binom{n}{y}}{(m+n+1)\binom{m+n}{z+y}}, \tag{4.24}$$

for $y = 0, 1, ..., n$. As an example, consider a consignment of size $N = 10$, where five units are inspected and all five are found to be illicit ($m = z = 5$). For the proportion of illicit units in the consignment to be at least 0.7 ($\theta \geq 0.7$), it is necessary for the number of units Y in the five units not inspected to be at least 2 ($Y \geq 2$). The beta-binomial probability in (4.24) with a uniform prior $f(\theta) = Be(1, 1)$ is given by

$$\Pr(Y \geq 2 \mid 5, 5, 5, 1, 1) = \sum_{y=2}^{5} \frac{6\binom{5}{5}\binom{5}{y}}{11\binom{10}{5+y}} = 0.985.$$

The beta-binomial approach enables a probability of 0.985 to be assigned to the event that $\theta \geq 0.7$, given a consignment of size 10 in which five units have been inspected and all found to contain drugs.

Compare this result with that obtained using a frequentist perspective based on the hypergeometric distribution (Sections 1.7.7 and A.2.5).

Examples of the use of the hypergeometric distribution for sampling in drugs related cases are given in Aitken (1999), Colón et al. (1993), Coulson et al. (2001b), Frank et al. (1991), Tzidony and Ravreboy (1992), Weusten (2011), and Moroni et al. (2012). The hypergeometric distribution is also recommended by the United Nations in this context (UN Office on Drugs and Crime, 2009). An application to fibres for the determination of the optimal sample size is given by Faber et al. (1999).

Let $R = Z + Y$ be the total number of units in the consignment that contain illicit drugs, where Z is the number of units in the sample of size m and Y is the number of units in the remainder which contain drugs. Then the distribution of Z is hypergeometric (Section A.2.5) with

$$\Pr(Z = z) = \frac{\binom{R}{z}\binom{N-R}{m-z}}{\binom{N}{m}},$$
$$(z = 0, 1, \ldots, \min(R, m)).$$

When $z = m$ this expression simplifies to

$$\Pr(Z = m) = \frac{R!(N-m)!}{N!(R-m)!}. \tag{4.25}$$

Consider an example where $z = m$ and $N = 10$. Probabilities in (4.25) are computed for some values of m (i.e. $m = \{4, 5, 6\}$) and for some values of θ (i.e. $\theta = \{0.6, 0.7\}$ so that $R = N\theta = \{6, 7\}$), with results in Table 4.3.

Table 4.3 Probabilities all m inspected units contain drugs for $m = 4, 5, 6$ in a population of size 10 where $\theta = 0.6, 0.7$ in a super-population.

θ	m		
	4	5	6
0.6	0.07	0.02	0.005
0.7	0.17	0.08	0.03

If $\theta = 0.7$, so when $N = 10$, $R = 7$, the probability all five inspected items out of 10 contain drugs is 0.08. Thus $\theta = 0.7$ is the 92% lower confidence bound for the proportion of illicit items in the super-population. If $\theta = 0.6$, so when $N = 10$, $R = 6$, the probability all five inspected items out of 10 contain drugs is 0.02. Thus $\theta = 0.6$ is the 98% lower confidence bound for the proportion of illicit items in the super-population. Similarly, for $N = 10$ and $m = 4$, if all 4 tablets sampled contain illicit drugs, then one can be 83% confident that the proportion of illicit drugs in the consignment is at least 0.7. Finally, for $N = 10$ and $m = 6$, if all 6 tablets sampled contain illicit drugs, then one can be 97% confident that the proportion of illicit items in the consignment is at least 0.7.

It can be observed that the beta-binomial and hypergeometric distributions give similar numerical answers, though Bayesian and frequentist approaches give different interpretations to the results (Aitken, 1999). The hypergeometric

distribution has the interpretation that if $m = z = 5$, one is 92% confident that $\theta \geq 0.7$. The beta-binomial approach enables one to assign a probability of 0.985 to the event that $\theta \geq 0.7$. As an aside, ignore the consignment size N. Assume a binomial distribution with $\theta = 0.6$. The probability a sample of size 4 would be all illicit is $0.6^4 = 0.13$. This is considerably different from the exact value of 0.07 provided by the hypergeometric distribution for a consignment size of 10.

General results can be obtained. The problem is to choose m such that, given n, α, and β (and possible values for z, consequential on the choice of m and on the outcome of the inspection), a value for y can be determined to satisfy some probabilistic criterion, e.g. the value y_0 such that $\Pr(Y \geq y_0 \mid m, n, \alpha, \beta) = p$. Results are given in Aitken (1999) for $p = 0.9$, where the consignment size N is taken to be 30.

If six units are inspected and one or two do not contain drugs then the number of units in the remainder of the consignment which can be said, with probability 0.9, to contain drugs drop from 17 to 12 to 9. Even if 16 units (out of 30) are inspected and all are found to contain drugs, then it can only be said, with probability 0.9, that 12 of the remaining 14 contain drugs (and this is so even with $\alpha = 4$, $\beta = 1$). See also Moroni et al. (2012) where a sequential approach is described.

These approaches for sample size estimation assume that the classification of items as licit

or illicit is free of error. It is obviously desirable that this assumption be true. An extension of the proposed Bayesian approach to account for possible laboratory errors has been proposed by Biedermann et al. (2008b) by means of graphical models and will be described briefly in Section 4.4.1. Further discussion of error is given in Zamengo et al. (2011).

An additional benefit of the assumption that analyses are error-free is that the posterior distribution of the proportion of the consignment, which is illicit is robust to the choice of the prior parameters. When there is a possibility of misclassification, the posterior distribution is no longer robust to the choice of the prior parameters. Such a situation is not discussed here but details are available in Rahne et al. (2000) A frequentist approach using the hypergeometric distribution with an adaptation to allow for false positives and false negatives is described in Faber et al. (1999). Application of these ideas to the sampling of glass fragments is described in Curran et al. (1998a).

There are instances where experiments give rise to more than two mutually exclusive events. In such a case, the binomial distribution can be generalised to the multinomial distribution and the beta distribution can be generalised to the Dirichlet distribution (Section A.3.8). A generalisation may be made, analogously, from the beta-binomial distribution to a so-called Dirichlet-multinomial distribution (Section A.2.7). An extension to sampling with a categorical response in which

there may be more than two possible responses (e.g. with pills, the responses may be LSD, ecstasy, and licit) is given in Mavridis and Aitken (2009).

4.4 BAYESIAN NETWORKS FOR SAMPLING INSPECTION

The question of sampling size can also be approached by graphical models, notably Bayesian networks. Bayesian networks allow for (i) a flexible analysis of sampling issues with the user being able to interact directly with possible models, (ii) an explicit and visual representation of underlying modelling assumptions, and (iii) calculation of posterior probability distributions for a consignment's true proportion of positives.

The core ideas related to Bayesian networks have been introduced in Chapter 2 (Section 2.9), and a wide description of the implementation of Bayesian networks in forensic science can be found in Taroni et al. (2014a).

4.4.1 Large Consignments

A Bayesian network for inference about the proportion of items showing a target characteristic of interest in a large consignment has been proposed by Biedermann et al. (2008b) and is depicted in Figure 4.10 (see also Taroni et al. (2014a, p. 304)). The proportion θ is modelled

with a discrete chance node named θ whose states represent disjoint intervals between 0 and 1 (this allows for an acceptable approximation of a continuous entity). The probability assigned to the intervals of node θ are determined by a beta distribution whose parameters α and β are provided by the nodes α and β, which are parents of node θ. Any values (> 0) can be defined to represent the analyst's prior beliefs about θ. A particular aspect of the Bayesian network shown in Figure 4.10 is the way in which the inspection procedure is modelled. Instead of a node for the overall number of successes (e.g. positive units) amongst the m that have been inspected, separate nodes are used to represent the target characteristic ('positive' or 'negative', 'licit' or 'illicit', and so on) of each inspected unit. The result of an examination or analysis on the i-th unit is modelled by a binary node *obs i* with states 'positive' and 'negative'.

Another interesting aspect of the Bayesian network in Figure 4.10 is that it accounts for the possibility of errors in the analysis. There may be circumstances that shed doubt on the result of an analysis. The condition of a unit chosen for inspection or erroneous experimental settings are possible reasons for this. Thus, an analysis may not always provide a positive result when the inspected unit is truly positive or may provide a positive result when the inspected unit is truly negative. For these reasons, a distinction is made between the true, but unknown, condition of a unit (e.g. containing or not containing an illegal

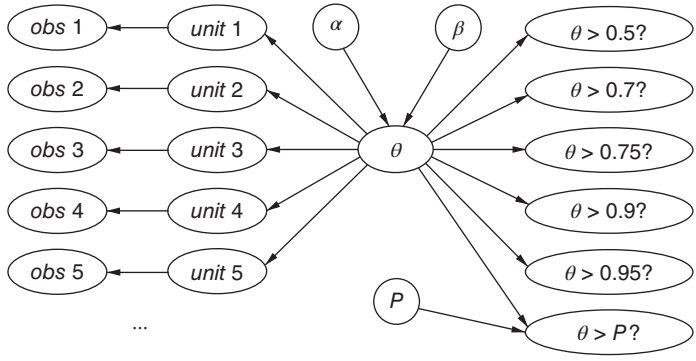

Figure 4.10 A Bayesian network for inference about a proportion of a large consignments. The definition of the nodes are as given in Table 4.4. Source: From Biedermann et al. (2008b). Reprinted with permissions of Oxford University Press.

substance) and what is observed in the course of an experiment designed to detect the presence or the absence of that target characteristic. Such distinctions are advocated, for example, in the context of DNA profiling analyses (Thompson et al., 2003), and are also challenged in other forensic disciplines (Saks and Koehler, 2005). Propositions according to which the *i*-th inspected unit may or may not contain an illegal substance is modelled by a binary node *unit i*. The outcome of a given analysis depends directly on the presence or the absence of the respective characteristic. Directed edges thus connect the nodes *obs i* and *unit i*. Generally, two probabilities can be used to describe the accuracy: the probability of a positive result when the unit is truly positive and the probability of a negative result when the unit is truly negative, the sensitivity and specificity,

respectively, (Section 4.2.3) of a test (e.g. (Balding, 2005; Kaye, 1987c; Lindley, 2006; Robertson and Vignaux, 1995b)). They can be used to complete the probability tables of the nodes *obs i*.

Before inspection of a unit, the probability that it is found to contain or not to contain an illegal substance depends directly on the proportion of units in the consignment that contain illegal substances. Directed edges are thus drawn from the node θ to the nodes *unit i*. The model also contains auxiliary nodes, namely, P, $\theta > 0.5?$, $\theta > 0.7?$, $\theta > 0.75?$, $\theta > 0.9?$, $\theta > 0.95?$, which define a substructure from which cumulative probabilities $\Pr(\theta > \theta_0 \mid \alpha, \beta, m, z)$ (4.11) may be evaluated.

The definition of the nodes is given in Table 4.4. Five pairs of nodes (*obs*, *unit*) are incorporated in the current model. Clearly, more trials may be needed for routine use and additional nodes can be added analogously. The proportion θ of positives in a consignment is modelled with a discrete chance node θ with intervals $0 - 0.05, 0.05 - 0.10, \ldots, 0.95 - 1$. The number of intervals as well as their size may be varied according to the analyst's needs.

Consider again a scenario in which $m = 4$ units are inspected and all are found to be positive. In such a scenario, as was illustrated in Section 4.3.1, there is – assuming a uniform prior probability for θ – a probability equal to 0.97 that the proportion of positive units θ is greater than 0.5 (see Table 4.1). The Bayesian network described earlier is depicted in Figure 4.11 with nodes expanded

Table 4.4 Definitions of nodes used in the Bayesian network shown in Figure 4.10.

Node	Definition	States
α, β	Parameters of the beta distribution defined for the node θ	$0.5, 1, 2, ..., 10$
θ	Proportion of positives in the consignment	$0 - 0.05, ..., 0.95 - 1$
P	Lower limit for the evaluation of cumulative probabilities of the proportion of positives in the consignment	$0, 0.05, ..., 0.95, 1$
$\theta > P$?	Is the proportion of the positives in the consignment greater than P?	*yes*, *no*
$\theta > 0.5$?	Is the proportion of the positives in the consignment greater than 0.5?	*yes*, *no*
(0.7, ...)	than 0.5 (0.7, ...)?	
obs 1 (2, ...)	Outcome of test conducted in order to determine the characteristic of item 1 (2, ...)	*positive*, *negative*
unit 1 (2, ...)	True (but unknown) characteristic of item 1 (2, ...)	*positive*, *negative*

and instantiations made at the relevant nodes. Nodes α and β are set to 1, whilst nodes *unit* 1 to *unit* 4 are set to *positive*. Instantiations of nodes *unit i* rather than *obs i* follows from the assumption that the determination of the characteristic of an inspected unit is made without error (see Taroni et al., (2010) for the examination of a setting in which the determination of the analytical characteristics cannot be assumed to be error free). The node $\theta > 0.5$? displays the target probability $\Pr(\theta > 0.5 \mid 1, 1, 4, 4)$. Notice the decrease in probabilities for $\Pr(\theta > 0.7), \Pr(\theta > 0.75)$, etc., as expected.

Notice further that consideration is not only limited in the proposed Bayesian network to the evaluation of findings that have actually been obtained. The probability may also be evaluated with which future trials, given previous observations, can be expected to yield positive and negative results, respectively. This is illustrated in Figure 4.11 where a probability of approximately 0.83 (0.832) is indicated for the fifth unit being positive.

4.4.2 Small Consignments

A Bayesian network able to be used for inference about the proportion of items showing a target characteristic of interest in a small consignment has been proposed by Biedermann et al. (2008b) and is depicted in Figure 4.12 (see also Taroni et al. (2010, pp. 267–270)). The target node of

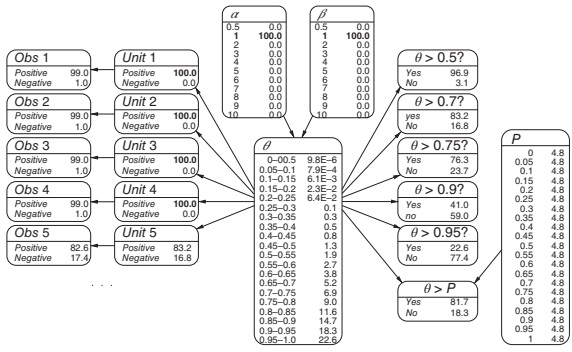

Figure 4.11 A Bayesian network for inference about a large consignment of discrete units. The definitions of the nodes are as given in Table 4.4. Four units are analysed and found to be positive (assuming error-free analyses). A uniform prior distribution is assumed for parameter θ.

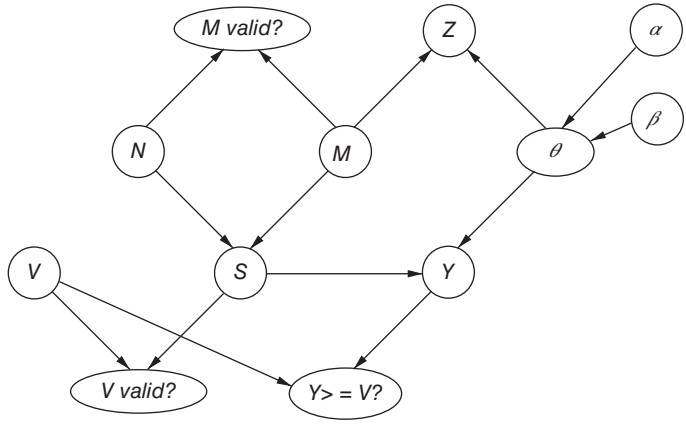

Figure 4.12 A Bayesian network for inference about a proportion of a small consignment. The definitions of the nodes are as given in Table 4.5.

the Bayesian network is Y, the number of positive units amongst those not analysed. The node definitions are given in Table 4.5,

The Bayesian network is structured as follows. Nodes N, M, and S are discrete chance nodes modelling the consignment size, the number of inspected items (sample size), and the number of uninspected items, respectively. The nodes Z and Y represent, respectively, the number of positives in the sample and the number of positives amongst the uninspected items. Following the model discussed in Section 4.3.2, the distribution of Y is binomial with parameters given by the nodes S and the node θ. The network also provides a substructure that assures that the instantiations that may be made at the nodes N and M satisfy the constraint $n \geq m$ (node M *valid*?), and that the

Table 4.5 Definitions of nodes used in the Bayesian networks shown in Figure 4.12.

Node	Definition	State
N	Consignment size	$n = 0, 1, ..., 20$
M	Sample size	$m = 0, 1, ..., 20$
S	Number of units not inspected	$s = 0, 1, ..., 20$
Z	Number of positive units in the sample	$z = 0, 1, ..., 20$
Y	Number of positives amongst the Uninspected units	$y = 0, 1, ..., 20$
V	Lower limit for evaluating cumulative probabilities of the number of positives amongst the uninspected units	$v = 0, 1, ..., 20$
$Y >= V$?	Are there at least V positives amongst the uninspected units?	*yes*, *no*
M *valid*?	Constraint on M	*true*, *false*
V *valid*?	Constraint on V	*true*, *false*
θ	Proportion of positives in the consignment	$0 - 0.05, ..., 0.95 - 1$
α, β	Parameters of the beta distribution defined for the node θ	$0.5, 1, 2, ..., 10$

instantiations that may be made at the nodes S and V satisfy the constraint $s \geq v$ (node V *valid*?). Finally, the node $Y \geq V$ provides the probability for the event that the number of positives amongst the units not inspected (Y) is at least equal to or greater than a specified number (V).

Consider again the scenario that was described in Section 4.3.2, where five units were inspected from a consignment of size 10 and all were found to be positive. Assuming a uniform prior distribution, a probability of 0.985 was found for the cumulative probability to find at least two positive units in the remainder of the consignment. The Bayesian network described earlier is depicted in Figure 4.13 with nodes expanded and instantiations made at the relevant nodes. Nodes α and β are set to 1, node N is set to 10, nodes M and Z are set to 5 whilst node V is set to 2. The node $Y \geq V$? displays the target probability $\Pr(Y \geq 2 \mid 5, 5, 5, 1, 1) = 0.985$.

4.5 INFERENCE FOR A NORMAL MEAN

Consider now data in the form of measurements, known as continuous data. For example, consider a scientist who is interested in determining the true level θ of alcohol concentration in the blood of someone suspected of driving under the influence of alcohol. The determination is to be done on the basis of a series of measurements (also known as

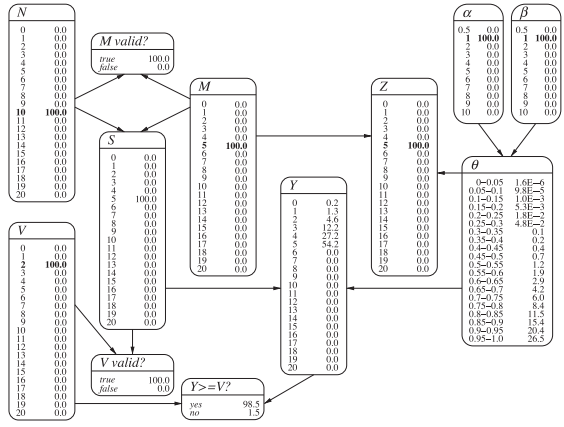

Figure 4.13 A Bayesian network for inference about a small consignment of discrete units. The definitions of the nodes are given as in Table 4.5. Five units are analysed from a consignment of size 10 and found to be positive. A uniform prior distribution is assumed for parameter θ. Source: Adapted from Taroni et al. (2010). ©John Wiley and Sons.

repeated measurements) taken by traffic police. The repeated measurements x may be considered equivalent in practice to the outcome of drawing a random sample from a Normally distributed random variable X, whose mean θ is the unknown quantity of interest.

Within the Bayesian paradigm, θ has a probability distribution and it is desired to determine this in order to make inferences about θ. The variance σ^2 of measurements may be treated as known or unknown. There are circumstances where it can be plausibly approximated from *ad hoc* calibrations (Howson and Urbach, 1996) and it is taken to be known. Inference about a normal mean when the variance is known is addressed in Section 4.5.1. However, there are circumstances where the variance cannot be taken to be a known quantity and it will be necessary to introduce a prior distribution also for this quantity. Inference about a normal mean where the variance is also unknown is addressed in Section 4.5.2.

4.5.1 Known Variance

Consider a random variable X that has a Normal distribution with mean θ and variance σ^2 (assumed known), so that

$$(X \mid \theta, \sigma^2) \sim N(\theta, \sigma^2).$$

The common choice of distribution for θ, the mean of a Normal distribution, is itself a Normal distribution. The Normal distribution is a conjugate prior

distribution for the mean of a Normal distribution. In the situation where the variance of the measurements is known, the prior distribution for θ and the distribution for $(X \mid \theta, \sigma^2)$ with known σ^2 combine together to give a posterior distribution, which is also a Normal distribution.

Denote the parameters of the Normal distribution for θ by ν, the mean, and τ^2, the variance. The distribution is represented by

$$(\theta \mid \nu, \tau^2) \sim N(\nu, \tau^2).$$

Then it can be shown (e.g. Lee (2012)) that the posterior distribution of θ, given a value x for X, and σ^2, ν, and τ^2 is

$$(\theta \mid x, \sigma^2, \nu, \tau^2) \sim N(\theta_1, \tau_1^2), \tag{4.26}$$

where

$$\theta_1 = \frac{\sigma^2 \nu + \tau^2 x}{\sigma^2 + \tau^2}, \tag{4.27}$$

and

$$(\tau_1^2)^{-1} = (\sigma^2)^{-1} + (\tau^2)^{-1}, \tag{4.28}$$

or, equivalently,

$$\tau_1^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}.$$

The posterior mean, θ_1, is a weighted average of the prior mean ν and the observation x, where the weights are the variance of the observation x and the variance of the prior mean ν, respectively, such that the component (observation or prior mean)

that has the smaller variance has the greater contribution to the posterior mean.

The reciprocal of the variance is known as the *precision*. Thus the precision of the posterior distribution of θ is the sum of the precisions of the prior distribution and the observation.

This result can be generalised to consider the distribution of the mean θ of a set of n independent, identically Normally distributed observations $x_1, \ldots, x_n$ with mean θ and variance σ^2. The generalisation follows from the result that $\bar{X}$, the random variable corresponding to the sample mean $\bar{x}$ of a sample of size n, has a distribution, known as the *sampling distribution*,

$$(\bar{X} \mid \theta, \sigma^2) \sim N(\theta, \sigma^2/n).$$

The posterior distribution of θ is Normally distributed with mean

$$\theta_1 = \frac{\frac{\sigma^2}{n}\theta + \tau^2\bar{x}}{\frac{\sigma^2}{n} + \tau^2}, \tag{4.29}$$

and variance

$$\tau_1^2 = \frac{\tau^2\sigma^2/n}{\sigma^2/n + \tau^2}, \tag{4.30}$$

or precision

$$(\tau_1^2)^{-1} = n(\sigma^2)^{-1} + (\tau^2)^{-1}. \tag{4.31}$$

The posterior mean θ_1 is a weighted average of the prior mean ν and the sample mean $\bar{x}$, with weights

proportional to the variances corresponding to the prior distribution and the sampling distribution. The mean of the distribution with lower variance (or higher precision) receives greater weight. The posterior precision is the sum of the precisions of the prior and the likelihood.

Consider the example of a person whose measurement x of the blood alcohol level is measured as 0.85 g/kg (see also Section A.3.2). It is of obvious interest to determine the probability that the true level θ of blood alcohol is greater than the legal threshold (say 0.80 g/kg). From (4.26), a posterior distribution for θ may be obtained, where the posterior mean θ_1 and the posterior variance τ_1^2 are computed following the updating rules given in (4.27) and (4.28). Application of these methods requires values to be given for the prior mean, ν, and variance, τ^2. The choice of these values may well be subjective and would have to be done carefully given the legal context in which the measurement x is being made. As illustrated in Section 4.1 there is considerable debate about the role of prior distributions in the law and forensic science. One approach that has been suggested to overcome the subjectivity of the choice of prior is the use of a vague prior, that is, a prior which provides poor information compared to the information provided by the data. When the unknown parameter lies in a finite interval, a uniform prior distribution may be applicable, that is, a distribution that is taken to be constant over the range of the variable of interest. This idea

was already introduced in Section 4.2, where the parameter θ of interest was a proportion.

In the context of the mean of a Normal distribution the range of interest is from $-\infty$ to $+\infty$. For the mean of a Normal distribution, it is not possible to take a constant value over the range of interest and retain the properties of a probability distribution as the probability density function will not integrate to 1. In such a circumstance, the prior distribution is then known as an *improper* or *vague* prior distribution. However, such a choice of prior may be acceptable if it combines with a likelihood to give a proper posterior distribution. In this example, the improper prior is uniform as it takes a constant value over the whole real line.

This is so for the Normal distribution. The uniform prior distribution is defined as the limiting distribution in which $\nu = 0$ and $\tau^2 \to \infty$. Inspection of (4.27) and (4.28) shows that the limiting values for the posterior mean and variance are simply $\theta_1 = x$ and $\tau_1^2 = \sigma^2$; i.e. for a uniform prior, the posterior distribution of θ is

$$\theta \sim N(x, \sigma^2). \tag{4.32}$$

In the blood alcohol example, given a measurement $x = 0.85$, a known variance for measurements from this procedure of 0.005, and a uniform (improper) prior for the true level θ, the posterior distribution of the true blood alcohol level, θ, is $N(0.85, 0.005)$. The probability that the true blood alcohol level is greater than a legal threshold

$\theta_0 = 0.8$ g/kg is then

$$\begin{aligned}
&\Pr(\theta > 0.8 \mid \theta \sim N(0.85, 0.005)) \\
&\quad = 1 - \Pr(\theta \leq 0.8 \mid \theta \sim N(0.85, 0.005)) \\
&\quad = 1 - \Phi\left(\frac{0.80 - 0.85}{\sqrt{0.005}}\right) \\
&\quad = 1 - \Phi(-0.7071) = 0.76,
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standardised normal distribution (Section A.3.2).

This result invites the question as to whether this reading of 0.85, combined with a uniform prior, is sufficient to find the suspect guilty beyond reasonable doubt of having a blood alcohol content greater than 0.80 g/kg. The assumption of a uniform prior is very supportive of the defence and implies the procedures are very imprecise, an inference that may be drawn from a value of $\tau^2 \to \infty$. The assumption may also be perceived to be attractive because it is generally felt, through the choice of such a distribution, that one is not committing oneself to any particular value. A uniform prior distribution can be interpreted to mean that no parameter value is favoured over any other. However, it is rarely the case that one has no information at all about the possible level of alcohol in blood and that any non-negative value can be considered as equally likely. The prior distribution can be determined from past experiments, the experience of the expert or from

the literature (e.g. values of a given magnitude are not physically justifiable). In the case at hand, suppose available knowledge (typically, circumstantial information, such as the fact that the person has been stopped by traffic police whilst driving erratically, exceeding the speed limit and so on) suggests a prior distribution centred around 1.3 with a standard deviation equal to 0.02, $\theta \sim \mathrm{N}(1.3, 0.0004)$. Note that this is equivalent to a judgement that *a priori*, a level of alcohol in blood lower than 1.24 and larger than 1.36 (i.e. more than three standard deviations from the mean) is considered very unlikely. A blood sample is analysed and a measurement $x = 0.85$ g/kg is obtained. The posterior distribution for θ from (4.27) and (4.28) is $\mathrm{N}(1.27, 0.0003)$. The 95% lower probability bound for θ is obtained from the 95% lower bound for a Normal distribution centred at 1.27 with variance equal to 0.0003. This is 1.24. There is a probability of 0.95 that the true value of θ is greater than 1.24.

Note also that if more than one measurement ($n > 1$) were taken, the inference for the true alcohol blood level would be based on the mean of the n measurements and the posterior parameters determined from (4.29) and (4.30). However, in the current example there is only one available measurement, and the variability σ^2 (0.005) of the procedure is large with respect to the prior uncertainty represented by τ^2 (0.0004). This will result in a posterior inference strongly effected by the prior distribution. Note that in any Bayesian

analysis, it is important to assess the sensitivity of any inferences with respect to changes in the model assumptions, including assumptions about the probability density $f(x \mid \theta)$ and the prior density $f(\theta)$. The Normal assumption is often made for computational convenience, and a sensitivity analysis might suggest consideration of a long-tailed alternatives such as the t-distribution (Section A.3.4) (Gelman et al., 2014). In what follows, a sensitivity analysis will be performed to explore the sensitivity of posterior inference about alcohol level concentration with respect to the choice of parameters in the prior distribution (i.e. the prior mean and the prior variance). The prior distribution may be felt too informative, with a high prior mean and a very small variance that makes the weight of the measurements very small. Figure 4.14 illustrates the impact of the choice of smaller values of ν and larger values of τ^2 on the posterior mean. It can be observed that as the prior variance increases relative to the variance of the observations then the posterior mean comes closer to the mean of the observations.

4.5.2 Unknown Variance

When the variance of the measurements cannot be assumed known, it is necessary to have an uninformative prior for it. A common such prior is a so-called Jeffreys' prior, Section A.3.3.

So far, the variance of the distribution of a random variable X has been assumed known. If

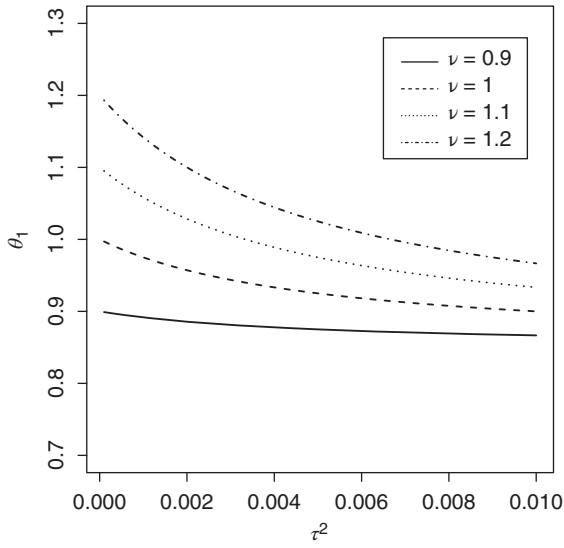

Figure 4.14 Posterior mean θ_1 of alcohol concentration (g/kg) as a function of the prior variance τ^2 for different levels of the prior mean: $\nu = 0.9$ (solid line), $\nu = 1.0$ (dashed line), $\nu = 1.1$ (dot-dashed line), $\nu = 1.2$ (dotted line), given a sample mean of 0.85 g/kg with variance $\sigma^2 = 0.005$.

this is not the case and it is necessary to consider a prior distribution for both parameters, then the posterior distribution for θ is related to the *t*-distribution (see Section A.3.4, Bernardo and Smith (2000); Bolstad and Curran (2017); Robert (2007); Lee (2012)). Again, whenever very little prior information is available (or it is desired to have a large consensus about the prior choice), Jeffreys' prior is appropriate.

Consider the prior mean first. The concept that only poor knowledge is available *a priori* (compared with the information the experiment is

expected to provide) can be expressed by assuming indifference between parameter values that lead to a likelihood which is completely determined *a priori*, except for its location. This may be expressed by adopting a uniform prior, that is, $f(\theta) = 1$, for $-\infty < \theta < \infty$. Consider now the prior variance, it can be shown that a uniform prior over σ^2 does not represent a locally uniform prior, whilst a uniform prior over $\log(\sigma^2)$ does. A vague prior distribution on σ^2 can be taken as $f(\sigma^2) = 1/\sigma^2$ (see e.g. (Lee, 2012); (Taroni et al., 2010)).

A non-informative prior distribution on both parameters is defined as the product of the two locally uniform priors now introduced and is given by

$$f(\theta, \sigma^2) = \frac{1}{\sigma^2}. \tag{4.33}$$

Suppose there are n observations of X, $x_1, \ldots, x_n$. The sample mean is $\bar{x} = \sum_{i=1}^{n} x_i$ and the sample variance is

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1).$$

The marginal posterior distribution of the population mean θ is now a non-central t-distribution with $(n-1)$ degrees of freedom, centred at $\bar{x}$ with spread parameter s^2/n (Section A.3.4). The probability density function of the transformed variable

$$t = \frac{(\theta - \bar{x})}{s/\sqrt{n}}.$$

is a (central) t-density with $(n-1)$ degrees of freedom and probabilistic inferences about θ can be made with reference to the t-distribution.

Subjective prior distributions can be elicited when more information is available. A conjugate prior distributions for (θ, σ^2) can be chosen as

$$f(\theta, \sigma^2) = f(\theta \mid \sigma^2) f(\sigma^2), \qquad (4.34)$$

where $\theta \mid \sigma^2$ is Normal with mean ν and variance σ^2/n_0 for some fixed prior parameters ν and n_0, and S_0/σ^2 is a chi-squared distribution with k degrees of freedom with fixed prior parameters S_0 and k (Section A.3.5). Notice that as the number of degrees of freedom increase, the probability distribution is more concentrated at smaller values of σ^2 (see Figure A.4 in Section (A.3.5)). Note that the prior distributions in (4.34) are not independent as the prior distribution about the mean θ depends on the population variance σ^2. It is possible for a joint prior distribution $f(\theta, \sigma^2)$ to be taken as the product of independent conjugate priors for each parameter, that is, $f(\theta, \sigma^2) = f(\theta) f(\sigma^2)$. If so, greater computations are required as the posterior distribution is not known in closed form and numerical solutions must be implemented. However, the choice to model the joint prior distribution $f(\theta, \sigma^2)$ as the product of dependent conjugate priors as in (4.34) can be justified not only on grounds of convenience. Since the prior distributions are often informed

by previous observations, it may be reasonable to calibrate prior beliefs about the population mean by the scale of measurements of the observations (Gelman et al., 2014). Parameter n_0 can be considered in terms of the prior sample size for the prior distribution for θ expressing the strength of belief about the chosen prior location ν. In other words, it can be interpreted as the sample size of Normally distributed observations that would have the same precision as the prior belief about θ (Bolstad and Curran, 2017). The marginal distribution of θ is a t-distribution with $k' = k + n$ degrees of freedom, centred at the mean θ_1 of the conditional posterior distribution of θ

$$\theta_1 = \frac{n\bar{x} + n_0\nu}{n + n_0},$$

with dispersion parameter equal to σ_1^2/n' where

$$\sigma_1^2 = \frac{S_0 + \sum_{i=1}^{n}(x_i - \bar{x})^2 + \left(\frac{n_0 n}{n_0+n}\right)(\bar{x} - \nu)^2}{k'}, \quad (4.35)$$

and $n' = n_0 + n$.

Consider again the scenario described in Section 4.5.1 where it was of interest to infer the quantity of alcohol in blood given available measurements that were assumed to be Normally distributed with known variance. Suppose a new procedure is available, so that the variability of the measurements cannot be treated as known. Two replicate measurements ($n = 2$) are available for this new

procedure, say $x_1 = 0.84$ and $x_2 = 0.86$. Consider first a non-informative prior distribution for (θ, σ^2) as in (4.33). From the available measurements x_1 and x_2, $\bar{x} = 0.85$ and $s^2 = 0.0002$. Then

$$t = \frac{(\theta - \bar{x})}{s/\sqrt{n}} = \frac{(\theta - 0.85)}{\sqrt{0.0002/2}} \tag{4.36}$$

has a t-distribution with 1 degree of freedom.

The 95% lower probability bound for θ is obtained from the 95% lower bound for a t-distribution with 1 degree of freedom. This is -6.31. Then, solution of

$$\frac{(\theta - 0.85)}{\sqrt{0.0002/2}} = -6.31$$

gives $\theta = 0.78$ There is a probability of 0.95 that the true value of θ is greater than 0.78.

Alternatively, an informative prior distribution as in (4.34) may be introduced. The prior distribution for θ given σ^2 can be centred at $\nu = 1.30$, as in Section 4.5.1, with $n_0 = 1$. As far as the prior distribution about σ^2 is concerned, consider a number of degrees of freedom $k = 1$ (see Figure A.4 in Section (A.3.5)). In this way the probability mass will be concentrated for smaller values of σ^2. Parameter S_0 can be elicited using available information about the variability of measurements of the old procedure, that was known and set equal to 0.005. One can consider that the new procedure will have better performance with respect to the older one, and that the variability

of measurements should be smaller. Suppose it is believed that $\Pr(\sigma^2 > 0.005) = 0.1$. Then

$$\Pr\left(\frac{\sigma^2}{S_0} > \frac{0.005}{S_0}\right) = \Pr\left(\frac{S_0}{\sigma^2} < \frac{S_0}{0.005}\right) = 0.1,$$

where S_0/σ^2 has a chi-squared distribution with k degrees of freedom and $S_0/0.005$ is therefore the 10% point of a chi-squared distribution with $k = 1$ degrees of freedom, that is 0.015 79, obtainable from statistical software. Therefore,

$$S_0 = 0.005 \times 0.015\ 79 = 0.000\ 078\ 9,$$

so σ^2 has $S_0 = 0.000\ 078\ 9$ times an inverse chi-squared distribution with 1 degree of freedom (Section A.3.6). The posterior distribution of the target quantity θ will be a t-distribution with $k' = 1 + 2 = 3$ degrees of freedom, centred at

$$\theta_1 = \frac{n\bar{x} + n_0\nu}{n + n_0} = \frac{2 \times 0.85 + 1 \times 1.30}{2 + 1} = 1,$$

and

$$\sigma_1^2 = \frac{S_0 + \sum_{i=1}^{n}(x_i - \bar{x})^2 + \left(\frac{n_0 n}{n_0 + n}\right)(\bar{x} - \nu)^2}{k'} = 0.045\ 09.$$

The dispersion parameter is therefore $\sigma_1^2/n' = 0.045\ 09/3 = 0.015\ 03$, which is set equal to $s/\sqrt{n}$. Then

$$t = \frac{(\theta - \bar{x})}{s/\sqrt{n}} = \frac{(\theta - 1)}{\sqrt{0.015\ 03}} \tag{4.37}$$

has a t-distribution with 3 degrees of freedom.

The 95% lower probability bound for θ is obtained from the 95% lower bound for a t-distribution with 3 degrees of freedom. This is -2.35. Then, solution of

$$\frac{(\theta - 1)}{\sqrt{0.015\ 03}} = -2.35$$

gives $\theta = 0.711$ There is a probability of 0.95 that the true value of θ is greater than 0.711.

4.5.3 Interval Estimation

Consider now interval estimation about the normal mean θ given the posterior distribution $N(\theta_1, \tau_1^2)$ in (4.27) and (4.28) (or in (4.29) and (4.30) when more than one observation is available). Because of the symmetry of the Normal distribution, the HPD interval for a given probability value is symmetric about the posterior mean, and it is equi-tailed. When the variance is known, the $100(1-\alpha)\%$ HPD interval is given by

$$[q_{\alpha/2}, q_{1-\alpha/2}],$$

where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are the $100\alpha/2\%$ and $100(1-\alpha/2)\%$ points of a Normal distribution $N(\theta_1, \tau_1^2)$ so that for $\theta \sim N(\theta_1, \tau_1^2)$

$$\Pr\{q_{\alpha/2} \leq \theta \leq q_{1-\alpha/2}\} = 1 - \alpha. \tag{4.38}$$

The HPD interval can also be obtained by rewriting (4.38) as

$$\Pr\left\{\frac{q_{\alpha/2} - \theta_1}{\tau_1} \leq \frac{\theta - \theta_1}{\tau_1} \leq \frac{q_{1-\alpha/2} - \theta_1}{\tau_1}\right\} = 1 - \alpha, \tag{4.39}$$

where $(q_{\alpha/2} - \theta_1)\tau_1^{-1} = -z_{1-\alpha/2}$ and $(q_{1-\alpha/2} - \theta_1)\tau_1^{-1} = z_{1-\alpha/2}$ are the $100\alpha/2\%$ and $100(1-\alpha/2)\%$ points of a standard Normal distribution (Section A.3.2). The $100(1-\alpha)\%$ HPD interval can therefore be obtained as

$$(\theta_1 - z_{1-\alpha/2}\tau_1, \theta_1 + z_{1-\alpha/2}\tau_1). \tag{4.40}$$

Consider again the context of the estimation of the concentration, θ, of alcohol in blood introduced earlier in Section 4.5.1. The measured alcohol concentration X in blood was assumed to be Normally distributed with variance known equal to 0.005, so $X \sim \mathrm{N}(\theta, 0.005)$, whilst both a non-informative and an informative prior distribution was considered for parameter θ. Consider the informative prior distribution $N(1.30, 0.0004)$. A blood sample is analysed and a measurement $x = 0.85$ g/kg is obtained. The posterior distribution for θ from (4.27) and (4.28) is $\mathrm{N}(1.27, 0.0003)$ and a 95% equi-tailed HPD interval is as follows:

$$(1.27 \pm 1.96\sqrt{0.0003}) = (1.23, 1.30). \tag{4.41}$$

The scientist is therefore entitled to say that their degree of belief that the true alcohol level is in fact in the realised interval $(1.23, 1.30)$ is equal to 0.95.

Compare the result in (4.41) with that obtained from a frequentist perspective. A confidence interval of level $(1-\alpha)$ for the Normal mean can be

obtained in the following way. Consider a Normal distribution $N(\theta, \sigma^2)$ for the quantity of interest X, then

$$\Pr\left\{-z_{1-\alpha/2} \leq \frac{X-\theta}{\sigma} \leq z_{1-\alpha/2}\right\} = 1-\alpha. \quad (4.42)$$

Given the available measurement x, the $100(1-\alpha)\%$ confidence interval for θ is

$$[x \pm z_{1-\alpha/2}\sigma].$$

In the case at hand, a 95% confidence interval can be obtained as

$$(0.85 \pm 1.96 \times 0.005^{1/2}) = (0.71, 0.99). \quad (4.43)$$

This interval is very different from (4.41) because no account is taken in (4.43) of prior information.

In many practical situations available measurements are more abundant ($n > 1$). Consider a random sample $(X_1, \ldots, X_n)$ from the distribution $X \sim \mathrm{N}(\theta, \sigma^2)$, then the result in (4.42) can be generalised as

$$\Pr\left\{-z_{1-\alpha/2} \leq \frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right\} = 1-\alpha,$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean and $\bar{X} \sim N(\theta, \sigma^2/n)$ (see Section A.3.2). Given a sample

of observations $x_1, \ldots, x_n$, the sample mean $\bar{x}$ is calculated. Then, the $100(1-\alpha/2)$ confidence interval for θ is

$$(\bar{x} \pm z_{1-\alpha/2}\sigma/\sqrt{n}).$$

Note that, whenever the population variance σ^2 is unknown, it is estimated by the sample variance s^2 and the quantiles $\pm z_{1-\alpha/2}$ will be substituted by the corresponding quantiles of a t distribution with $(n-1)$ degrees of freedom.

The conflict between the Bayesian credible interval in (4.41), $(1.23, 1.30)$, and the confidence interval in (4.43), $(0.71, 0.99)$, is not surprising. Recall results in (4.27) and (4.29) according to which the posterior mean is a weighted average of the prior mean and observations with weights given by the population variance (i.e. 0.005 or $0.005/n$) and the prior variance (i.e. 0.0003). Since there is only one measurement available, and the population variance is higher than the prior variance, the posterior distribution of the true quantity of alcohol θ is dominated by the prior distribution. The reader can verify that whenever a uniform prior distribution is taken for θ (as in Section 4.5.1), the probability interval and the confidence interval coincide. However, the interpretation differs since the philosophies underlying their respective constructions are different.

4.6 QUANTITY ESTIMATION

Consider a setting where it is of interest to estimate the quantity of drugs in a consignment of packets. It is only possible to make a statement about the consignment as a whole with certainty if the whole consignment is analysed (and no error is committed). Once it is accepted that a sample has to be considered, it is necessary to consider what level of proof is adequate. This is strictly a matter for the court to decide.

This section addresses and answers the following question:

- Given a sample from a consignment of homogeneous material, what is the quantity of illicit material in the consignment?

A frequentist approach based on Student's *t*-distribution is described in Mario (2010) and Alberink et al. (2014, 2017). However it is a Bayesian approach that is discussed here.

A probability interval is appropriate. In a Bayesian context, a probability distribution is associated with a parameter (Q, say) denoting the total quantity of illicit material in the consignment and probability statements of any desired kind may be made. For example, these could include the probability that Q is greater than a certain value, which will be of importance in sentencing hearings.

The estimation of the quantity of drugs will be treated in two stages. First the proportion of the units in the consignment that contain illicit drugs will be modelled. Secondly, the total weight of the illicit material in those packets that do contain anything illicit is estimated. Consider the consignment as itself a random sample from a large super-population of units or packages, some or all of which contain illegal material. Then θ $(0 < \theta < 1)$ is the proportion of units in the super-population that contain illegal material. Uncertainty about the proportion of packets that are illicit may be represented by a beta distribution (Section A.3.7).

Consider a setting in which a consignment of $N = m + n$ units is seized. A number (m) of units are examined; the choice of m may be made following the procedures described in Section 4.3. On examination it is found that z $(\leq m)$ units contain drugs and that $(m - z)$ do not. The contents of the z units that contain drugs are weighed and their weights $(x_1, \ldots, x_z)$ recorded. The remainder $(n = N - m)$ are not examined. All of m, z, and n are known. Let $y(\leq n)$ be the number of units that contain drugs amongst the units that are not examined. Clearly, y is unknown. Let $(w_1, \ldots, w_y)$ be measurements of the quantity of drugs in those units not examined which contain drugs. Let $\bar{x} = \sum_{i=1}^{z} x_i/z$ be the sample mean quantity of drugs in units amongst those examined and found to contain drugs, and let s be the sample standard deviation where the sample variance

$s^2 = \sum_{i=1}^{z} (x_i - \bar{x})^2/(z-1)$. Let $\bar{w} = \sum_{j=1}^{y} w_j/y$ be the mean quantity of drugs in units containing drugs amongst those not examined. Clearly, $\bar{w}$ is unknown. The total quantity q of drugs in the exhibits is then $(z\bar{x} + y\bar{w})$ and the problem is one of first estimating $\bar{w}$, given $\bar{x}$, s, and z, whilst not knowing y and then of finding $y\bar{w}$ by finding the posterior distribution of $Y \mid \bar{x}$. An *estimative* approach is one in which the parameters (θ, σ^2) of the Normal distribution representing the quantity of drugs in an individual unit are estimated by the corresponding sample mean $\bar{x}$ and sample variance s^2 (Tzidony and Ravreboy, 1992). A *predictive* approach is one in which the values of the unknown measurements $(w_1, \ldots, w_y)$ are *predicted* by values of known measurements (Aitchison and Dunsmore, 1975; Aitchison et al., 1977; Evett et al., 1987; Geisser, 1993).

The predictive approach *predicts* the values of $\bar{w}$ (and hence q) from $\bar{x}$ and s through the probability density function $f(\bar{w} \mid \bar{x}, s)$

$$f(\bar{w} \mid \bar{x}, s) = \int f(\bar{w} \mid \theta, \sigma^2) f(\theta, \sigma^2 \mid \bar{x}, s) d\theta \, d\sigma^2,$$

where $f(\theta, \sigma^2 \mid \bar{x}, s)$ is the posterior density function for (θ, σ^2) based on a prior density function $f(\theta, \sigma^2)$ and the summary statistics $\bar{x}$ and s. When prior information for θ and σ^2 is not available a uniform prior for θ and σ^2 may be used as in Section 4.5.2. The predictive density function $f(y \mid \bar{x})$ is then a generalised t-distribution as described in Sections 4.6.1 and 4.6.2.

The advantage of the predictive approach relative to the estimative approach is that any prior knowledge of the variability in the parameters (μ, σ^2) of the Normal distribution can be modelled explicitly. Suggestions as to how this may be done are given by Aitken et al. (1997) with reference to *U.S v. Pirre* (1991).

4.6.1 Predictive Approach in Small Consignments

The probability density function $f(q)$ of Q has been derived, for small and large consignments, respectively, in Aitken and Lucy (2002). First, consider a small consignment. Let Y $(\leq n)$ denote the number of units not examined, which contain drugs. The estimation of quantity is able to take account of the lack of knowledge of Y. A probability function for Y may be determined using the methods described in Section 4.3. A weighted average of the quantities obtained for each value of Y is taken with weights the probabilities of Y obtained from an appropriate beta-binomial distribution (Section A.2.7). Let $(X_1, \ldots, X_z)$ and $(W_1, \ldots, W_Y)$ be the weights of the contents of the units examined and not examined, respectively, which contain drugs. It is assumed that these weights are Normally distributed. Let $(x_1, \ldots, x_z)$ be the observed values of $(X_1, \ldots, X_z)$. The total weight, Q, of the contents of the units in the consignment is then given by

$$Q = z\bar{x} + Y\bar{W},$$

where $\bar{x} = \sum_{i=1}^{z} x_i/z$ and $\bar{W} = \sum_{j=1}^{Y} W_j/Y$. The distribution of $(Q \mid x_1, \ldots, x_z)$, which is a predictive distribution, is of interest. Once known, it is possible to make probabilistic statements about Q.

Let $\bar{x}$ and $s = \sqrt{\sum_{i=1}^{z} (x_i - \bar{x})^2/(z-1)}$ denote the mean and the standard deviation, respectively, of the measurements on the z units, which were examined and found to contain drugs. The number of units not examined equals n, of which y (unknown) contain drugs and for which the mean quantity of drugs is $\bar{w} = \sum_{j=1}^{y} w_j/y$ (and it is unknown).

A lower bound q for the quantity Q of drugs can be derived from the relationship $Q = z\bar{x} + Y\bar{W}$ as follows. First, condition on $Y = y$. Then

$$\begin{aligned}
&\Pr(Q < q \mid m, z, n, y, \bar{x}, s) \\
&\quad = \Pr(z\bar{x} + y\bar{W} < q \mid m, z, n, y, \bar{x}, s) \\
&\quad = \Pr\left(\bar{W} < \frac{q - z\bar{x}}{y} \mid m, z, n, y, \bar{x}, s\right). \qquad (4.44)
\end{aligned}$$

In the absence of prior information about the mean or the variance of the distribution of the weights of drugs in the packages, a uniform prior distribution is used. Details of how such prior information may be considered is given in Aitken et al. (1997) (for statistical considerations) and in Bring and Aitken (1997) (for legal considerations). Given $Y = y$, the quantity

$$\frac{\bar{W} - \bar{x}}{s\sqrt{\frac{1}{z} + \frac{1}{y}}}$$

has a t-distribution with $(z-1)$ degrees of freedom (Section A.3.4). Quantiles of this distribution and hence lower bounds for the quantity $q = z\bar{x} + y\bar{w}$, according to appropriate burdens of proof may be determined.

For given values of m, z, n, y, $\bar{x}$, and s, lower bounds for $\bar{w}$ and hence q can be determined from the formula

$$\bar{w} = \bar{x} + s\, t_{z-1,\alpha} \sqrt{\frac{1}{z} + \frac{1}{y}},$$

where $t_{z-1,\alpha}$ is the quantile of order α of a t-distribution with $(z-1)$ degrees of freedom. Let $T = (\bar{W} - \bar{x})/s\sqrt{(1/z) + (1/y)}$. Then, from (4.44) one obtains

$$\Pr\left(T < \frac{q - (z+y)\bar{x}}{sy\sqrt{\frac{1}{z} + \frac{1}{y}}} \mid m, z, n, y, \bar{x}, s \right),$$

where T has a t-distribution with $z-1$ degrees of freedom. Let

$$t_{qy} = \frac{q - (z+y)\bar{x}}{sy\sqrt{\frac{1}{z} + \frac{1}{y}}}.$$

For a small consignment, the value of y is a realization of a random variable, which has a beta-binomial distribution (Section A.2.7). The conditional distribution of $\bar{W}$, given $Y = y$, $(\bar{W} \mid m, z, n, y, \bar{x}, s)$, can be combined with the marginal distribution of Y to give a distribution $(\bar{W} \mid m, z, \bar{x}, s)$. The distribution and corresponding probability function of Q may then be determined from the

relationship $Q = z\bar{x} + y\bar{W}$, say

$$\Pr(Q < q \mid m, z, n, y, \bar{x}, s) = \sum_{y=0}^{n} \Pr(T < t_{qy} \mid m, z, n, y, \bar{x}, s) \Pr(Y = y).$$

The probability density function $f(q)$ of Q can be derived by differentiation of the cumulative distribution function. Let $f_{t,z-1}(\cdot)$ denote the probability density function of the t-distribution with $(z-1)$ degrees of freedom. The probability density function $f(q)$ of Q is then given by

$$f(q) = \sum_{y=0}^{n} f_{t,z-1}(t_{qy}) \left(sy \sqrt{\frac{1}{z} + \frac{1}{y}} \right)^{-1} \Pr(Y = y). \tag{4.45}$$

The values for Q corresponding to appropriate percentage points of the distribution may be determined from (4.45). Some results are given in Table 4.6 together with frequentist lower bounds using the fpc factor (4.15) and in Figure 4.15.

Estimative Approach

Given the sample size, and thus an estimate of the proportion of a consignment which contain drugs, and an estimate of the mean and the standard deviation of the weight in the consignment, a confidence interval for the true quantity of drugs may be calculated following the method described by Tzidony and Ravreboy (1992), according to which the consignment is considered as a

Table 4.6 Estimates of quantities q of drugs, in a consignment of $m + n$ units, according to various possible burdens of proof, expressed as percentages $P = 100 \times \Pr(Q > q \mid m, z, n, \bar{x}, s)$ in 26 packages when 6 packages are examined ($m = 6$, $n = 20$) and $z = 6$, 5, or 4 are found to contain drugs.

Percentage P	Number of units examined that contain drugs			Possible burden of proof (illustrative)
	6	5	4	
97.5	0.689 (0.930)	0.501 (0.744)	0.345 (0.575)	
95	0.750 (0.968)	0.559 (0.785)	0.397 (0.613)	Beyond reasonable doubt
70	0.944 (1.067)	0.770 (0.885)	0.603 (0.704)	Clear and convincing
50	1.015 (1.105)	0.862 (0.921)	0.704 (0.737)	Balance of probabilities

The mean ($\bar{x}$) and standard deviation (s) of the quantities found in the packages examined which contain drugs are 0.0425 and 0.0073 g. The parameters for the beta prior are $\alpha = \beta = 1$. Numbers in brackets are the corresponding frequentist lower bounds using the fpc factor (4.15).
Source: From Aitken and Lucy (2002). Reprinted with permissions of ASTM International.

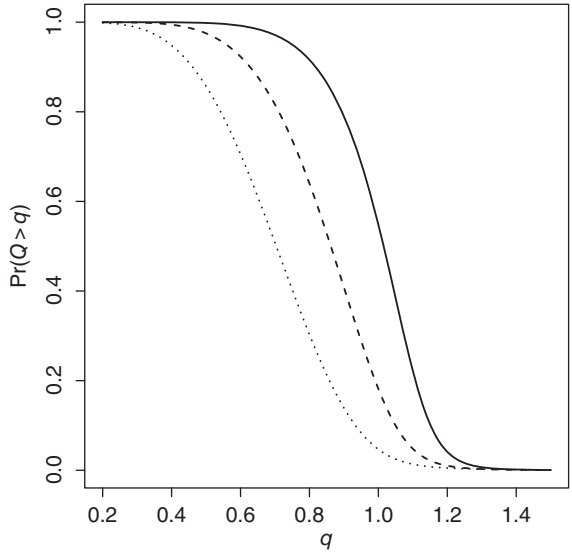

Figure 4.15 The probability that the total quantity Q of drugs (in grams) in a consignment of 26 packages is greater than q when 6 packages are examined and 6 (solid curve), 5 (dashed), or 4 (dot-dashed) are found to contain drugs. The mean and standard deviation of the quantities found in the packages examined which contain drugs are 0.0425 and 0.0073 g. The parameters for the beta prior are $\alpha = \beta = 1$. Source: Reproduced with permission from Aitken and Lucy (2002). ©ASTM International.

population and the packages (or units) examined as a sample. The quantities (weights) of drugs in the units are assumed to be random variables that are Normally distributed, with population mean θ and population variance σ^2, say. The mean quantity in a unit in the consignment is estimated by the mean, denoted $\bar{x}$, of the quantities of found in the sample. A confidence interval is determined

for θ based on the sample size m, the sample mean $\bar{x}$, the sample standard deviation s of the quantities of drugs in the unit examined, and an associated t-distribution. Following results in Tzidony and Ravreboy (1992), a $100(1-\alpha)\%$ confidence interval for the mean quantity in a package results

$$\left(\bar{x} \pm t_{m-1,1-\alpha/2}\frac{s}{\sqrt{m}}\sqrt{\frac{N-m}{N}}\right),$$

where $t_{m-1,1-\alpha/2}$ is the $100(1-\alpha/2)\%$ point of a t-distribution with $(m-1)$ degrees of freedom, and $\sqrt{(N-m)/N}$ is the finite population corrector factor. An estimate of the total quantity of drugs in the consignment is then determined by considering the size N of the consignment and the proportion θ of packages in the consignment thought to contain drugs. The corresponding confidence interval for Q, the total quantity of drugs in the consignment is obtained by multiplying the lower and the upper bound of the interval by $N\hat{\theta}$ where $\hat{\theta}$ is an estimate for θ based on the sample size m:

$$N\hat{\theta}\left(\bar{x} \pm t_{m-1,1-\alpha/2}\frac{s}{\sqrt{m}}\sqrt{\frac{N-m}{N}}\right).$$

However, no account is taken of the uncertainty in the estimation of θ, only a point estimate of θ is used.

An example in which a seized drug exhibit contained 26 street doses is given in Tzidony

and Ravreboy (1992). A sample of six ($m = 6$) units was taken and each was weighed. Twenty ($n = 20$) units were not examined. All six of the units examined contained drugs. The average net weight $\bar{x}$ of the powder in the six units was 0.0425 *g* with a standard deviation *s* of 0.0073 *g*. A 95% confidence interval for the total quantity in the 26 doses is

$$26[0.0425 \pm 2.57 \times 0.0073/\sqrt{6} \times \sqrt{20/26}] = (0.93, 1.28).$$

Note that this interval incorporates the finite population correction factor to allow for the relatively large sample size ($m = 6$) compared with the consignment size ($N = 26$). The Bayesian approach described earlier does not require such a correction.

It is also possible to consider just one end of the confidence interval, a so-called *confidence limit*. A lower limit on the quantity of drugs is desired as it enables the court to determine a limit at the appropriate level of proof above which the true quantity lies. For example, from Frank et al. (1991) a statement of the form that 'at a 95% confidence level, 90% or more of the packages in an exhibit contain the substance' is suggested by them as being sufficient proof in cases of drug handling that 90% or more of the packages contain the substance. The nature of the construction of the confidence limit is that in 95% of cases studied for which these results are obtained then 90% of

the packages will contain the substance. However, the probability with which a particular interval in a particular case contains the true proportion is not known as there is no randomness associated to the proportion. From a frequentist perspective, the proportion is an unknown fixed quantity, and the realised confidence interval will either contain it or not. A corresponding $100(1-\alpha)\%$ lower bound for Q is given by

$$N\hat{\theta}\left(\bar{x} - t_{m-1,1-\alpha}\frac{s}{\sqrt{m}}\sqrt{\frac{N-m}{N}}\right).$$

The lower end 0.93 g of the 95% confidence interval (0.93, 1.28) g for the quantity Q of drugs in the 26 packages may be thought of as an approximate 97.5% lower confidence limit for Q. This can be compared with the value 0.689 g in the corresponding cell of Table 4.6, which is the amount such that $\Pr(Q > 0.689) = 0.975$ obtained from the predictive approach. The predictive approach produces a lower value because of the uncertainty associated with the values determined for the number of unexamined units that contain drugs. This difference is repeated for different probabilities. In general, the Bayesian approach gives smaller values for the quantities than the frequentist approach.

Further details are available in Aitken et al. (1997), Izenman (2001), and Aitken and Lucy (2002) where it is shown that as the burden of proof, concerning the amount of drugs in

the packages, increases, the quantity for which charges may be brought decreases thus lowering the length of any sentence which may be related to quantity. For example, if proof is required *beyond reasonable doubt* and a probability of 0.95 is thought to meet this burden then the quantity associated with this is 0.750 g (assuming all six units examined contain drugs) since, from Table 4.6, $\Pr(Q > 0.750) = 0.95$. Alternatively, if proof is required on *the balance of probabilities* and a probability of 0.50 is thought to satisfy this, then the quantity associated with this is 1.015 g since, again from Table 4.6, $\Pr(Q > 1.015) = 0.50$. If less than six of the units examined are found to contain drugs then the estimates for q decreases considerably as can be seen from the second and the third columns of Table 4.6.

4.6.2 Predictive Approach in Large Consignments

For a large consignment the data are used to provide a beta posterior distribution for the proportion of illicit drugs in the whole consignment. It is assumed that the consignment size is known. The total weight, Q, of the contents of the units in the consignment is given as before as

$$Q = z\bar{x} + y\bar{W}.$$

The distribution of Q is then given by the t-density, conditional on y, with $\Pr(Y = y)$ replaced by an

appropriate part of a beta distribution over the interval $(0, n)$ (see Appendix A). Results for a large consignment, obtained by scaling up by a factor of 100 from the results in Table 4.6, are shown in Table 4.7 and Figure 4.16 with a similar pattern of results to those for small consignments. Note that in the t-density component of the expression y is treated as a discrete variable in the interval $\{0, \ldots, n\}$ and in the beta component of the expression it is treated as a continuous variable. The treatment of y as a continuous variable for

Table 4.7 Estimates of quantities q *g*. of drugs, in a consignment of $m + n$ units, according to various possible burdens of proof, expressed as percentages $P = 100 \times Pr(Q > q \mid m, z, n, \bar{x}, s)$ in 2600 packages when 6 packages are examined ($m = 6$, $n = 2594$) and $z = 6$, 5, or 4 are found to contain drugs.

Percentage P	Number of units examined that contain drugs		
	6	5	4
97.5	63 (95)	44 (78)	30 (61)
95	69 (98)	51 (80)	36 (63)
70	91 (106)	74 (88)	58 (70)
50	98 (110)	84 (92)	69 (74)

The mean ($\bar{x}$) and standard deviation (s) of the quantities found in the packages examined which contain drugs are 0.0425 and 0.0073 g. The parameters for the beta prior are $\alpha = \beta = 1$. Numbers in brackets are the corresponding frequentist lower bounds without using the fpc factor.
Source: From Aitken and Lucy (2002). Reprinted with permissions of ASTM International.

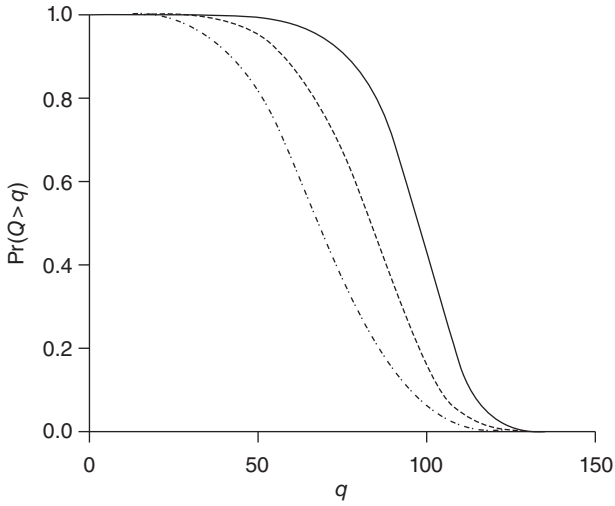

Figure 4.16 The probability that the total quantity Q of drugs (in grams) in a consignment of 2600 packages is greater than q when 6 packages are examined and 6 (solid curve), 5 (dashed), or 4 (dot-dashed) are found to contain drugs The mean and standard deviation of the quantities found in the packages examined that contain drugs are 0.0425 and 0.0073 g, respectively. The parameters for the beta prior are $\alpha = \beta = 1$. Source: From Aitken and Lucy (2002). Reprinted with permissions of ASTM International.

the beta integral enables the calculation of the probability that y takes a particular integer value for use with the t-density.

For more complicated cases see Alberink et al. (2014, 2017). A guideline is proposed for the estimation of the total amount of drugs in cases where there are items of different weights containing drugs in concentrations to which there is considerable measurement uncertainty attached.

Alberink et al. (2014) used *t*-distributions to determine confidence intervals for quantities that take account of variation in drug concentration amongst items and of measurement uncertainty with Normally distributed variation and constant relative standard deviations. Alberink et al. (2017) extended this work to examples that do not assume constant relative standard deviations.

4.7 DECISION ANALYSIS

The framework that was introduced in Section 2.8 applies well to problems of statistical inference described so far in this chapter, where the decision problem is the choice of a proper summary (e.g. an estimate) of a quantity of interest θ, and the decision space is the set Θ of possible θ values. According to a decision-theoretic perspective, a point estimate $\tilde{\theta}$ of a parameter θ represents a *decision* to act as though $\tilde{\theta}$ were the true value of θ. A loss function $L(\tilde{\theta}, \theta)$[2] thus measures the consequence of acting as if the true value of the quantity of interest were $\tilde{\theta}$ when it is actually θ, whilst the probability distribution $f(\theta)$ describes personal beliefs of a scientist about parameter θ (e.g. a Normal mean). To find the optimal decision, one must compute the Bayesian posterior expected loss

[2] $L(\tilde{\theta}, \theta)$ is used instead of the more correct $L(c(\tilde{\theta}, \theta))$ for ease of notation.

of a decision that averages the loss according to the posterior probability distribution $f(\theta \mid x)$ derived after observing measurements x that encapsulate all the available information about the parameter of interest at the time of decision making:

$$\mathrm{EL}(\tilde{\theta} \mid x) = \int_{\Theta} \mathrm{L}(\tilde{\theta}, \theta) f(\theta \mid x) d\theta. \tag{4.46}$$

An optimal decision, also called a *Bayesian decision*, is a decision θ^* that minimises the Bayesian expected loss in (4.46). This is the equivalent of (2.19) for a continuous parameter.

Such a decision framework can be extended to any kind of statistical inference, including interval estimation and hypothesis testing. An application to the problem of sampling is developed in Section 4.7.2. Further examples can be found in Taroni et al. (2010).

4.7.1 Standard Loss Functions

As for the prior distribution, a decision maker is allowed to choose any loss function that reflects their preferences, in particular the undesirability of alternative consequences they may face. However, the choice is best made with standard mathematically tractable loss functions. Consider the *squared-error* (or *quadratic*) loss:

$$\mathrm{L}(\tilde{\theta}, \theta) = k(\tilde{\theta} - \theta)^2 \tag{4.47}$$

where k denotes a constant (Press, 2003). The loss associated with making a decision $\tilde{\theta}$ when the

true state of nature is θ increases by the square of the difference between $\tilde{\theta}$ and θ (i.e. decisions $\tilde{\theta}$ far away from the true state of nature θ are strongly penalised). For example, if the difference doubles the loss quadruples. A squared-error loss with $k = 4$ and $\tilde{\theta} = 0.5$ is shown in Figure 4.17a. The Bayesian posterior expected loss is given by

$$\mathrm{EL}(\tilde{\theta} \mid x) = \int_{\Theta} k(\tilde{\theta} - \theta)^2 f(\theta \mid x) d\theta. \qquad (4.48)$$

It can be shown (see, e.g. (Berger, 1985)) that the Bayes decision θ^* under a quadratic loss, that is, the one minimising (4.48), is the posterior mean. The choice of a quadratic loss function may be acceptable whenever it is reasonable to equally penalise under-estimation and over-estimation, though this can be achieved by means of alternative symmetric loss functions. Consider for the sake of illustration the case where it is of interest to estimate the height of an individual appearing in video recordings made by a surveillance camera during a bank robbery (Taroni et al., 2006). In such a context it may be reasonable to accept that under- and over-estimation of the actual height will incur equal losses, and therefore a squared-error loss may be chosen (Taroni et al., 2010).

Another standard loss function is the *piecewise linear loss function*

$$\mathrm{L}(\tilde{\theta}, \theta) = \begin{cases} k_1(\theta - \tilde{\theta}) \text{ if } \theta - \tilde{\theta} \geq 0, \\ k_2(\tilde{\theta} - \theta) \text{ if } \theta - \tilde{\theta} \leq 0. \end{cases} \qquad (4.49)$$

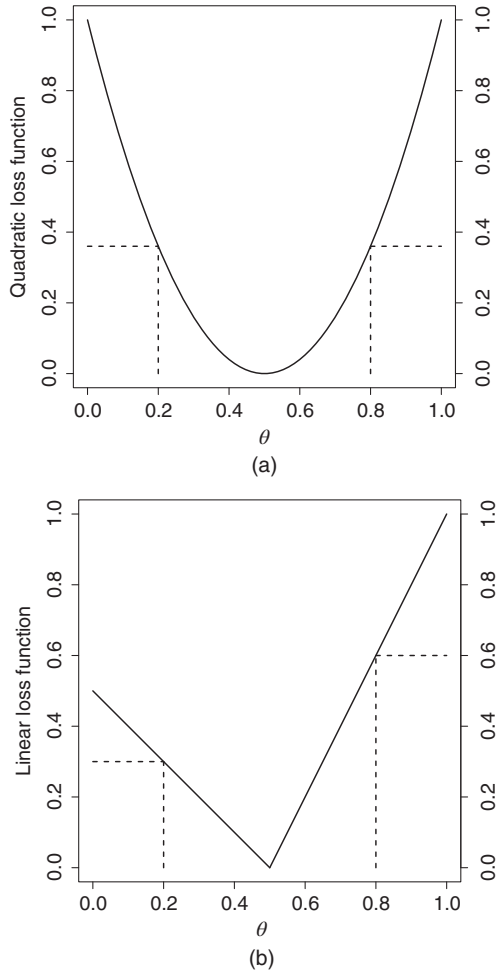

Figure 4.17 (a) Quadratic loss function in (4.47) with $\tilde{\theta} = 0.5$ and $k = 4$; (b) Piecewise linear loss function in (4.49) with $\tilde{\theta} = 0.5$, $k_1 = 2$, and $k_2 = 1$.

This loss increases more slowly than the quadratic loss and does not overpenalise large but unlikely errors. The Bayesian posterior expected loss is given by

$$\mathrm{EL}(\tilde{\theta} \mid x) = \int_{\theta>\tilde{\theta}} k_1(\theta - \tilde{\theta}) f(\theta \mid x) d\theta + \int_{\theta<\tilde{\theta}} k_2(\tilde{\theta} - \theta) f(\theta \mid x) d\theta. \quad (4.50)$$

It can be shown (see, e.g. (Berger, 1985)) that the Bayesian decision under a linear piecewise loss function is the $100k_1/(k_1 + k_2)\%$ point of the posterior distribution of the parameter of interest. The constants k_1 and k_2 can be chosen so as to reflect the relative importance of under-estimation and over-estimation. In the particular case where $k_1 = k_2$, the loss function is symmetric and the Bayes decision θ^* is the posterior median. A piece-wise linear loss function is illustrated in Figure 4.17b that penalises under-estimation more heavily ($k_1 > k_2$). One can easily observe that with values of θ equidistant from $\tilde{\theta}$ (in this case, $\tilde{\theta} = 0.5$), losses associated to an under-estimation are larger than those associated to an over-estimation of the same magnitude (e.g. acting as $\tilde{\theta} = 0.5$ when the true value θ_0 of θ is 0.8 is felt as more undesirable than acting as $\tilde{\theta} = 0.5$ when the true value of θ is 0.2.). Take, for example, the problem of estimating blood alcohol concentration in blood that was discussed earlier in Section 4.5, where – starting from a uniform

distribution – the posterior distribution for the true level θ of alcohol in blood was $N(0.85, 0.005)$. A decision maker may prefer to penalise the under-estimation of the true alcohol concentration more than the over-estimation. This may be so because falsely concluding a modest alcohol concentration in an individual with high alcohol concentration is regarded as a more serious error (because such an individual may represent a serious danger in traffic) than falsely assigning a high blood alcohol concentration to an individual which has actually a low concentration level. Given a piecewise linear loss function with $k_1 = 2$ and $k_2 = 1$, the Bayes decision θ^* is the $100(2/3)\%$ point of the posterior distribution, $\theta^* = 0.88$ g/kg. An asymmetric piecewise linear loss function was proposed by Taroni et al. (2014b) in forensic toxicology (analysis of the presence of THC in blood from car drivers).

As well as the choice of the prior distribution, a recurrent question is how to choose an appropriate loss function and particularly how this may impact on the optimal decision. A sensitivity analysis is performed to different choices of values k_1 and k_2 in (4.49). Results, in terms of the $100k_1/(k_1 + k_2)\%$ point of the posterior distribution (the value minimising the expected loss (4.50)), are given in Table 4.8.

The optimal decision is concerned with the Bayesian estimate of the quantity of interest. A different loss function may produce a different Bayes decision, but does not necessarily influence

Table 4.8 Sensitivity of the optimal decision on the alcohol level (g/kg) to different choices of k_1 and k_2 in the piecewise linear loss function in (4.49).

k_1	k_2				
	1	2	3	4	5
1	0.85	0.82	0.80	0.79	0.78
2	0.88	0.85	0.83	0.82	0.81
3	0.90	0.87	0.85	0.84	0.83
4	0.91	0.88	0.86	0.85	0.84
5	0.92	0.89	0.87	0.86	0.85

the decision to be taken by the recipient of the expert evidence (e.g. a change from a guilty to a non-guilty verdict in a legal contest). There are instances in Table 4.8 where the Bayesian estimate is lower than the threshold θ_0 (0.8), though these values are obtained for choices of k_2 much larger than k_1 (say, $k_2 = 4$ and $k_1 = 1$). These values would imply an unlikely preference structure according to which falsely concluding a modest alcohol concentration in an individual with high alcohol concentration is considered less severe than falsely assigning a high blood alcohol concentration to an individual which has actually a low concentration level.

Another standard loss function is related to a two-action decision problem. Suppose that the parameter space Θ can be partitioned into two non-overlapping sets Θ_1 and Θ_2 such that $\Theta = \Theta_1 \cup \Theta_2$. A question that may be of interest

is whether the true value of θ belongs to Θ_1 or to Θ_2, that is, to compare alternative hypotheses $H_1 : \theta \in \Theta_1$ and $H_2 : \theta \in \Theta_2$. The decision space can be described as $\mathcal{D} = \{d_1, d_2\}$. The first decision, d_1, amounts to accepting the view according to which the parameter θ takes values in Θ_1. The second decision, d_2, amounts to accepting the view according to which the parameter θ takes values in Θ_2. Formally, the two decisions d_1 and d_2 can be conceptualised as decisions to accept one of the two composite hypotheses: H_1 and H_2. A loss function for such a two-action decision problem can be defined (for $i = 1, 2$) as:

$$\mathrm{L}(d_i, \theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_i, \\ f_i(\theta) & \text{if } \theta \notin \Theta_i. \end{cases} \tag{4.51}$$

A decision d_i is accurate if the true value of parameter θ lies in the range defined by hypothesis H_i, and a zero loss is associated. Alternatively, incorrect decisions (i.e. when the true value of θ lies outside the range defined by hypothesis H_i) have an associated positive loss $f_i(\theta)$, where $f_i(\theta)$ is a function (could be linear, quadratic, or something else) of the distance between the decision d_i and the true value of θ. In other words, the loss depends on the severity of the mistake.

4.7.2 Decision Analysis for Forensic Sampling

Bayesian approaches illustrated in Section 4.3 consider sampling essentially as a problem of

the derivation of probabilistic statements about the composition of a consignment in order to answer the questions of interest as 'how big a sample should be taken' or 'what proportion of a consignment of discrete, homogeneous items is illicit'? Although the handling of the uncertainty through probability is an essential aspect of sampling scenarios, the situation that is actually faced by a customer of forensic expertise is one that contains elements that allow the outcome to be considered as a *problem of decision making*. This section will examine this point in further detail.

Consider the consignment inspection scenario that was described in Section 4.3, where a sample of size m from a large consignment (so that sampling can be taken as with replacement) is inspected, z items are found to be positive and the uncertainty about the proportion θ of items presenting a given target characteristic is modelled by a beta distribution with parameters α and β, $Be(\alpha, \beta)$. Two decisions might be detected. The first one, d_1, amounts to accepting the view according to which the proportion θ of 'positive' units (i.e. units with illicit content) in the consignment is greater than some specified value θ_0, say, $\theta > \theta_0$. The second one, d_2, is the view according to which the proportion θ of positive units in the consignment is not greater than the specified value θ_0, say, $\theta \leq \theta_0$. Note that in this case $\Theta_1 = (\theta_0, 1)$ and $\Theta_2 = (0, \theta_0]$.

A loss function as defined in (4.51), with $f_i(\theta) = k_i \mid \theta - \theta_0 \mid$, may be considered

$$\mathrm{L}(d_i, \theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_i, \\ k_i(\theta - \theta_0) & \text{if } \theta \notin \Theta_i \text{ and } \theta \geq \theta_0, \\ k_i(\theta_0 - \theta) & \text{if } \theta \notin \Theta_i \text{ and } \theta \leq \theta_0. \end{cases} \tag{4.52}$$

Given the stated loss function, the Bayesian posterior expected loss for d_1, that is, accepting $H_1 : \theta > \theta_0$ is

$$\begin{aligned} \mathrm{EL}(d_1 \mid z, m) &= \int_{\Theta_2} k_1(\theta_0 - \theta) f(\theta \mid z, m) d\theta \\ &= \int_{\Theta_2} k_1 \theta_0 f(\theta \mid z, m) d\theta \\ &\quad - \int_{\Theta_2} k_1 \theta f(\theta \mid z, m) d\theta, \end{aligned}$$

where $f(\theta \mid z, m) = Be(\alpha + z, \beta + m - z)$. Similarly, the Bayesian posterior expected loss for d_2, that is, accepting $H_2 : \theta \leq \theta_0$, is

$$\begin{aligned} \mathrm{EL}(d_2 \mid z, m) &= \int_{\Theta_1} k_2(\theta - \theta_0) f(\theta \mid z, m) d\theta \\ &= \int_{\Theta_1} k_2 \theta f(\theta \mid z, m) d\theta \\ &\quad - \int_{\Theta_1} k_2 \theta_0 f(\theta \mid z, m) d\theta. \end{aligned}$$

After some algebra it can be shown that (Taroni et al., 2010, pp. 214–215)

$$\mathrm{EL}(d_1 \mid z, m) = k_1 \left\{ \theta_0 \int_{\Theta_2} f(\theta \mid z, m) d\theta - \frac{\alpha + z}{\alpha + \beta + m} \int_{\Theta_2} f_1(\theta \mid z, m) d\theta \right\}, \tag{4.53}$$

and

$$\mathrm{EL}(d_2 \mid z, m) = k_2 \left\{ -\theta_0 \int_{\Theta_1} f(\theta \mid z, m) d\theta + \frac{\alpha + z}{\alpha + \beta + m} \int_{\Theta_1} f_1(\theta \mid z, m) d\theta \right\}, \tag{4.54}$$

where $f_1(\theta \mid z, m) = Be(\alpha' = \alpha + z + 1, \beta' = \beta + m - z)$.

Consider a specified value $\theta_0 = 0.80$. In Section 4.3.1 it has been shown that if a sample of size $m = 13$ is taken and all items turn out to be positive, and assuming a uniform prior distribution for θ, then the probability that the proportion of positive items is greater than 0.8 is approximately equal to 0.95, that is,

$$\int_{\Theta_1} f(\theta \mid z, m) d\theta = \Pr(\theta > \theta_0 \mid \alpha' = 1 + 13, \beta' = 1) = 0.95,$$

and therefore $\int_{\Theta_2} f(\theta \mid z, m) d\theta = 0.05$. The other probabilities in (4.53) and (4.54) can be obtained

analogously and

$$\begin{aligned}\int_{\Theta_1} f_1(\theta \mid z, m)d\theta \\ &= \Pr(\theta > \theta_0 \mid \alpha' = 1 + 13 + 1, \beta' = 1) \\ &= 0.964,\end{aligned}$$

and $\int_{\Theta_2} f_1(\theta \mid z, m)d\theta = 0.036$. Assume that values for k_1 and k_2 in (4.52) are taken equal to 1. The Bayesian posterior expected losses in (4.53) and (4.54) can be obtained as

$$\begin{aligned}\mathrm{EL}(d_1 \mid z, m) &= 0.8 \times 0.05 - \frac{1 + 13}{1 + 1 + 13} \times 0.035 \\ &= 0.0023,\end{aligned}$$

and

$$\begin{aligned}\mathrm{EL}(d_2 \mid z, m) &= -0.8 \times 0.95 + \frac{1 + 13}{1 + 1 + 13} \\ &\quad \times 0.964 = 0.135.\end{aligned}$$

The optimal decision d^* is the one that minimises the expected losses $\mathrm{EL}(d_i \mid z, m)$ in (4.53) and (4.54). Therefore, one should decide d_1 whenever $\mathrm{EL}(d_1 \mid z, m) < \mathrm{EL}(d_2 \mid z, m)$. The optimal decision here is d_1, since it minimises the loss.

One might object that equal values for k_1 and k_2 are not reasonable (or not justifiable). A false consideration that $\theta > \theta_0$ (i.e. decision d_1 is taken and is incorrect) may be felt as more undesirable than falsely considering $\theta < \theta_0$ (i.e. decision d_2 is taken and is incorrect). This would suggest taking

$k_1 > k_2$. It can be verified that, whenever 13 items are sampled and all are found to contain drugs (and a uniform prior distribution is assumed), decision d_1 will remain the optimal decision unless $k_1 > 60k_2$, that is, there is good reason to accept the idea that falsely considering $\theta > \theta_0$ is roughly 60 times worse than falsely considering $\theta < \theta_0$. It is possible to verify that whenever $k_1 = 60$ and $k_2 = 1$, the expected losses in (4.53) and (4.54) would become equal to 0.14 and 0.135, respectively.

The loss function in (4.51) can also take the form of a so-called '$0 - k_i$' loss function, where $f_i(\theta)$ is constant and equal to k_i, that is,

$$\mathrm{L}(d_i, \theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_i, \\ k_i & \text{if } \theta \notin \Theta_i. \end{cases}$$

In the consignment inspection scenario described so far, the loss k_1 associated with erroneously deciding d_1 could represent the amount of compensation to be allocated to an erroneously pursued individual, or the net loss represented by money that has been seized in a non-priority case (i.e. one in which it was considered, falsely, that $\theta > \theta_0$). In turn, the loss k_2 associated with erroneously deciding d_2 could consist of the funds or monetary value of property that could have been confiscated by the investigative authority as a penalty, and given to the public treasury. The Bayesian posterior expected loss in (4.46) is

given by

$$\begin{aligned}\mathrm{EL}(d_1 \mid z, m) &= k_1 \int_{\Theta_2} f(\theta \mid z, m) d\theta \\ &= k_1 \Pr(\theta \leq \theta_0 \mid \alpha' = \alpha + z, \beta' = \beta + m - z),\end{aligned} \tag{4.55}$$

and

$$\begin{aligned}\mathrm{EL}(d_2 \mid z, m) &= k_2 \int_{\Theta_1} f(\theta \mid z, m) d\theta \\ &= k_2 \Pr(\theta > \theta_0 \mid \alpha' = \alpha + z, \beta' = \beta + m - z).\end{aligned} \tag{4.56}$$

The optimal decision d^* is the one that minimises the expected losses $\mathrm{EL}(d_i \mid z, m)$ in (4.55) and (4.56). Therefore, one should decide d_1 whenever $\mathrm{EL}(d_1 \mid z, m) < \mathrm{EL}(d_2 \mid z, m)$. By rearranging terms, decision d_1 should be taken when

$$\Pr(\theta \leq \theta_0 \mid \alpha' = \alpha + z, \beta' = \beta + n - z) < \frac{k_2}{k_1 + k_2}.$$

As an example, consider $k_1 = k_2 = 100$, but readers may choose their own values, including of course asymmetric values $k_1 \neq k_2$. Recall the posterior distribution of θ is $Be(14, 1)$ (a uniform prior distribution was chosen, and a sample of size $m = 13$ has been inspected and all items are found to be positive), then

$$\Pr(\theta \leq 0.8 \mid 14, 1) = 0.04 < 1/2 \Rightarrow d^* = d_1.$$

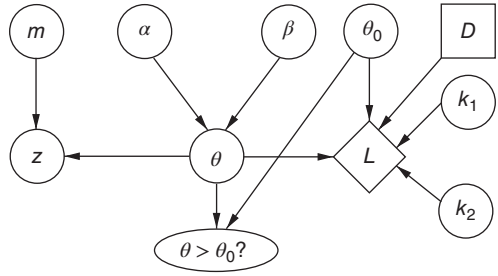

Figure 4.18 Collapsed representation of an influence diagram for deciding about a proportion when m units are inspected. The node θ represents the unknown proportion, the diamond shaped node L the loss and the squared node D the available actions. The remaining nodes are defined according to Table 4.9. Source: From Biedermann et al. (2012a). Reprinted with permissions of Elsevier.

An influence diagram (Figure 4.18) can be constructed to assist in the calculation of the posterior probabilities for θ and the associated expected losses for decisions d_1 and d_2 (see Section 2.9 and (Biedermann et al., 2012a)). The definitions of the nodes are summarised in Table 4.9. A network fragment covering discrete chance nodes α, β, θ, $\theta > \theta_0?$ follows the definitions given earlier in Section 4.4.1 (Table 4.4). The node $\theta > \theta_0?$ represents probabilities for values of the proportion that are greater than a specified value θ_0. This model fragment is extended with two discrete chance nodes m and z providing the number of inspected items and the number of items that are found to be positive (i.e. presenting the target characteristic), a decision node D (covering the two available decisions d_1 and d_2),

Table 4.9 Definitions of the nodes used in the influence diagram shown in Figure 4.18. The states of the nodes $\alpha, \beta, \theta_0, k_1$ and k_2 are chosen according to the requirements of the case under investigation. Other ranges of values may be chosen as required.

Node	Description	State(s)
m	Number of inspected items	$0, 1, 2, \cdots$
z	Number of inspected items found to be positive	$0, 1, \ldots$
α, β	Parameters of the beta distributed variable θ	$0.1, 0.5, 1, 2, ..., 10$ (e.g.)
θ	Proportion θ of 'positive' units in the population (i.e. seizure or consignment)	$0 - 0.05, ..., 0.95 - 1$
θ_0	Target value for the unknown proportion	$0.85, 0.9, 0.95$ (e.g.)
$\theta \geq \theta_0$?	Is the true proportion θ greater than the specified target value θ_0?	*true*, *false*
D	Decision about the proportion (available actions)	d_1, d_2
k_1	Loss for erroneously deciding $d_1 : \theta > \theta_0$ when $\Theta_2 : \theta \leq \theta_0$ is the true state of affairs	100 000 (e.g.)
k_2	Loss for erroneously deciding $d_2 : \theta \leq \theta_0$ when $\Theta_1 : \theta > \theta_0$ is the true state of affairs	100 000 (e.g.)
L	Decision loss	$L(d_i, \Theta_j),\ i, j = 1, 2$

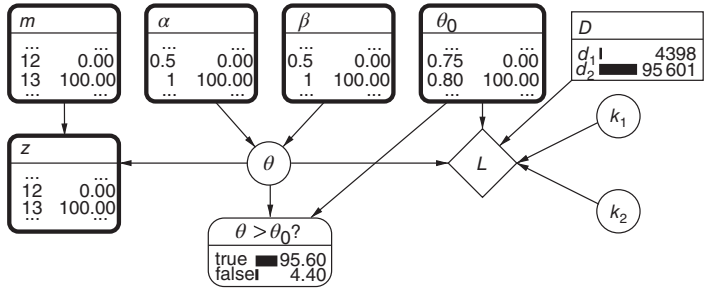

Figure 4.19 Partially expanded representation of the influence diagram in Figure 4.18 for deciding about a proportion when m units are inspected and $z = m$ are found to be positive. The node θ represents the unknown proportion, the diamond shaped node L the loss and the squared node D the available actions. The remaining nodes are defined according to Table 4.9. Instantiated nodes are shown with a bold border.

a loss node L for the decision consequences, and two discrete chance nodes k_1 and k_2 providing the losses. Node L takes the value 0, or the numerical values defined for k_1 and k_2, depending on the actual configuration of the parental nodes of L, that is decisions d_i, the true state of nature Θ_j, and the specified limiting value θ_0. The expanded representation of this Bayesian decision network is represented in Figure 4.19, where nodes m and z are instantiated to the value 13, to reflect available information. The node θ will account for the posterior distribution, based on a uniform prior distribution, whilst the posterior probability that the proportion θ is greater than a given threshold θ_0 (here θ_0 is set equal to 0.8) is provided by the node $\theta > \theta_0$?. The posterior expected losses

of decisions d_1 and d_2 are given in node D where $\text{EL}(d_1) = 4398$ and $\text{EL}(d_2) = 95\ 601$.

Consignment inspection can be expensive and it may be of interest to take into consideration also the cost of inspection. The Bayesian decision network in Figure 4.18 can be easily modified to take into account the cost of inspection (Biedermann et al., 2012a).

Another question that is regularly encountered in practice is whether it is worth inspecting individual items of a consignment or preferable to make a decision about the proportion θ without sample information. Aspects about such a so-called *pre-posterior* analysis are developed in Biedermann et al. (2012a).

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

5

Evidence and Propositions: Theory

5.1 THE CHOICE OF PROPOSITIONS AND PRE-ASSESSMENT

In 1998, Evett and Weir emphasised that the likelihood ratio represents the best available model for the scientist to understand and assess the value of an item of evidence. They expressed three principles (see Section 5.2), two of which are of primary importance for this chapter. These two principles stipulate that in order to evaluate any item of evidence, it is necessary to consider at least two propositions and to condition one's assessment on a framework of task-relevant circumstances. The perception of the circumstances has a clear impact on the choice of the propositions, which can broadly be organised within a so-called

hierarchical model. The nature of the propositions that the scientist will help address is therefore of particular interest. This chapter will present principles of interpretation and a methodology for the choice of relevant propositions. The formal development of the likelihood ratio expression at different hierarchical levels is explained.

Later in the book, in Chapter 6, the evaluation of evidence for different evidential types (e.g. fibres, DNA) is described in detail through practical examples. The evaluation of the value of recovered glass fragments is the forerunner of many of the ideas applied to other evidential types. These include, for example, firearms and toolmarks, fingermarks, speaker recognition, hair, documents and handwriting, and paint. Transfer material (such as fibres or glass fragments) is discussed in various different chapters throughout the book as it is one of the better evidential types for the discussion of the different ideas. There are many other evidential types for which the logic of evidence evaluation is not yet so well developed, for example, numerical (i.e. digital) evidence for which Biedermann and Vuille (2016) proposed ideas for its evaluation.

In this chapter, various general principles are presented that form the foundation of the statistical approach for evidence evaluation in various areas such as glass and fibres. Reference is also made to these areas and others (various transfer material, fingermarks and shoemarks)

elsewhere as appropriate, mainly in Chapter 6 where numerical examples are developed to illustrate the approach.

5.2 LEVELS OF PROPOSITIONS AND ROLES OF THE FORENSIC SCIENTIST

It is widely accepted that for the assessment of scientific evidence, the scientist should consider different propositions proposed by the prosecution and the defence to illustrate their description of the facts under examination.[1] The approach to this aspect of evidence evaluation was the subject of a large-scale project at the Forensic Science Service in the United Kingdom called the *Case Assessment Initiative*. The pioneering works are those of Cook et al. (1998a,b); and Evett et al. (2000e). This methodology is also described in one of a series of guides for judges, lawyers, forensic scientists, and expert witnesses published by the Royal Statistical Society (Jackson et al., 2014).

[1]The reason for the use of the term 'propositions' instead of 'hypotheses' is re-emphasised. Gittelson et al. (2016a) noticed that '[. . .] in science, hypotheses refer to statements that can be tested by performing a scientific experiment, whereas a 'proposition' is not necessarily testable. In forensic science, a statement such as 'The defendant murdered the victim' cannot be tested in a laboratory, so it is preferable to call this a 'proposition', and to refer generally to the statements corresponding to H_p and H_d as 'propositions'.' (p. 186)

Broadly speaking, propositions are statements of issues of dispute. Forensic scientists evaluate their findings under these propositions. The formulation of propositions is a crucial basis for a logical and scientific approach to the evaluation of evidence (e.g. Cook et al., 1998b; Buckleton et al., 2014). The framing of the propositions is an important and difficult stage of the evaluation process, because it depends on case information and the allegations of each of the parties, parties that have different key issues. The key issues are formally defined by the ENFSI Guideline (ENFSI, 2015):

> *The key issue(s) represent those aspects of a case on which a Court, under the law of the case, seeks to reach a judgement. The key issue(s) provide the general framework within which requests to forensic practitioners and propositions (for evaluative reporting) are formally defined. (p. 21)*

More generally, the framing of propositions can be guided by three key principles (Evett and Weir, 1998):

(1) Evaluation is only meaningful when at least two competing propositions are considered, conventionally denoted throughout this book as H_p and H_d, or occasionally, for example, $H_{d_i}(i = 1, \ldots, n)$.
(2) Evaluation of scientific evidence (E) considers the probability of the evidence given the propositions of interest, $\Pr(E \mid H_p, I)$ and

$\Pr(E \mid H_d, I)$, where I denotes the task-relevant information (see also point 3).

(3) Evaluation of scientific evidence is carried out within a framework of circumstances, denoted I. The evaluation is conditioned not only by the competing propositions but also by the structure and content of the framework.

Note that the principles were previously given in Evett (1990).

The importance of propositions in this evaluative framework is discussed in Taroni et al. (2013), Hicks et al. (2016), and Gittelson et al. (2016a).

Generally, propositions are considered in pairs. There will be situations where there will be three or more propositions and comments on these situations are given in Sections 2.3.2.2 and 6.1.6.2. This can happen with DNA mixtures, for example, where the number of contributors to the mixture is in dispute (Buckleton et al., 1998; Lauritzen and Mortera, 2002; Biedermann et al., 2011d; Buckleton et al., 2019) or in cases involving the DNA profile of a single stain of body fluids when relatives of the defendant are considered as potential donors of the recovered stain. Multiple propositions may also be of interest in cases of gunshot residue (GSR) when alternative actions are emphasised by the prosecution or the defence (e.g. 'The defendant shot the victim' versus 'The defendant did not shoot, but was a bystander'). It is generally possible with some thought to reduce the number of propositions to two, which will

be identified with the prosecution and defence positions, respectively.

Clearly the two propositions must be mutually exclusive. It is tempting to argue that they ought to be exhaustive, but this is not necessary (see Section 2.1.1). A simple way to achieve exhaustiveness is by adding the word 'not' in the first proposition. For example, changing 'Mr C is the man who kicked Mr Z' to 'Mr C is not the man who kicked Mr Z'. This, however, gives the court no idea of the way in which the scientist has assessed the evidence with regard to the second proposition. Mr C may not have kicked the victim, but he may have been present during the incident (e.g. Mr C may say that he helped the victim after the attack). Analogously, consider the proposition 'Mr B had sexual intercourse with Miss Y', modified to 'Mr B did not have sexual intercourse with Miss Y'. In fact, if semen has been found on the vaginal swab, then it may be inferred that someone has had sexual intercourse with Miss Y and, indeed, the typing results from the semen would be evaluated by considering their probability given that it came from some other man. It will help the court if this is made clear in the alternative proposition that is specified. The alternative could thus be 'Some unknown man, unrelated to Mr B, had sexual intercourse with Miss Y' (with no consideration of relatives). In summary, the simple use of 'not' to frame the alternative proposition is unlikely to be particularly helpful to the court (Cook et al., 1998b).

In the same sense, it is useful to avoid words such as 'contact' to describe the type of action in the main propositions of interest, because vague words may be misleading. As noted by Evett et al. (2000e), the statement that a PoI has been in recent contact with broken glass could mean many things. There is a need, thus, to specify propositions clearly in a framework of circumstances. Moreover, the scientist may confuse *propositions* with *explanations*. For example, statements like 'the crime stain originated from the PoI' and 'the crime stain originated from some unknown person who happened to have the same genotype as the PoI' represent explanations. The probability of the evidence that the DNA profile of the crime stain genotype matches the profile of the PoI given the first explanation is 1, but the probability of the evidence given the alternative explanation is also 1. Thus, the likelihood ratio is 1 and would not help advance resolution of the issues in the case. Analogously, if the propositions are defined as 'The matching DNA came from the defendant' and 'The matching DNA came from an unknown (unrelated) individual', the likelihood ratio equals 1, too (Hicks et al., 2015). The interpretation is uninformative in this case because the alternative proposition explains, so to say, the observation but does not enable the value of the evidence to be determined. Explanations can be useful as an exploratory tool and they play an important role in reconstructions that normally contribute to the investigative phase. Details and

examples of the distinction are presented by Evett et al. (2000e) and by Jackson et al. (2014).

It is also common to observe that results (i.e. observations on the recovered material) are included in the formulation of the propositions. This happens, for example, when the scientist, upon observing a given feature in the trace (e.g. a Nike logo in a shoemark), suggests that the alternative proposition is that an unknown Nike shoe left the mark. Hicks et al. (2015) explain, through various examples, why this should not be done. The main reason is that including observations in the propositions will mean that it will be for the court to evaluate the results, without benefiting from any advice from the scientist. So, if forensic scientists wish to assess the value of the evidence in a meaningful way, they should ensure that results do not overlap with the definition of propositions. On the practical side, the propositions that are of interest in a judicial case depend on (i) the circumstances of the case, (ii) the available background data, and (iii) the key issue(s) (criteria supported by ENFSI (2015)). A classification (a so-called *hierarchy*) of these propositions into three main categories or levels has been proposed, notably source level (level I), activity level (level II), and offence level (level III) propositions (Cook et al., 1998a).

It is important to note that – as presented by Buckleton et al. (2014) – in some cases, when there are no clear propositions or even no PoI, a forensic scientist may be able to generate explanations to account for the observations. In such a

case, the forensic scientist plays the role of *investigator* rather than that of *evaluator*. It is important to distinguish between the two roles and be aware of them. A discussion on the roles and on the inferential processes to guide thinking and practice in investigation and in court proceedings was presented in Jackson et al. (2006).

Bayes' factor (or likelihood ratio) can be used for two main purposes in forensic science:

(1) The first purpose consists of assigning a value for a given item of evidence. This refers to the evaluative level at which forensic science operates. Evaluation of a piece of evidence means that the scientist provides an expression of the value of the evidence in support – which may be positive, negative, or neutral – of a proposition of interest. This represents the task introduced in Section 2.3. An important aspect of this level of operation is that the scientist does not express an opinion about a proposition itself. This is the main difference with respect to the second purpose.

(2) The second purpose is the provision of information to the police. Here, the scientist acts at an investigative level. At this stage, the scientist tries to answer questions such as 'what happened?'. The forensic scientist is said to be 'crime focused' and observes evidence that forms the basis for the generation of propositions and suggestions for explanations, in order to give guidance to the police investigators.

As previously reported by Buckleton et al. (2014), in forensic settings it may be the case that a PoI is not available for comparative purposes. Therefore, it will not be possible to evaluate the characteristics observed in the recovered material and those in the control material as would be the case in an evaluative setting. Notwithstanding this, the data – measurements made on the recovered material – can be used for an investigative purpose. Scientists can offer to the police information in support of more general propositions. See Taroni et al. (2012b) for an example of questioned documents where the (investigative) propositions were 'the author of the questioned document (say, a threatening letter) is a male (or a female)' or 'the author of the questioned document is (is not) left-handed.

The assessment of the evidence at level I propositions in the hierarchy (i.e. source level) depends, in essence, on analyses and measurements on the recovered and control materials. The value of a trace (or, a mark, signal, or stain) given source level propositions (e.g. 'Mr. X's pullover is the source of the recovered fibres' and 'An unknown garment is the source of the recovered fibres') does not need to take into account anything else other than the analytical information obtained during examination. Aspects such as the presence or absence of material, or where a stain was found, are not considered. Further, the nature of the stain (e.g. a biological fluid or cells) is known. The probability of the evidence under the first proposition,

providing the numerator of the likelihood ratio, is considered from a careful comparison between two materials (i.e. the recovered and the control). The probability of the evidence under the second proposition, providing the denominator of the likelihood ratio, is based on a consideration of the comparison result, that is, the observed features, with respect to a population of alternative sources.

There can be uncertainty, however, regarding the relevance of the evidence. Because of the sensitivity of DNA current profiling technology, it is now possible to encounter situations in which it is not necessarily the case that a particular profile actually came from what was observed as a discernible region of staining. In such cases, it may be necessary to address what are termed 'sub-level I' propositions or sub-source propositions. In a DNA context, for example, level I propositions such as 'The semen came from Mr Smith' and 'The semen came from some other man' are replaced by 'The DNA came from Mr Smith' and 'DNA came from some other person', respectively. In that context, Taylor et al. (2016a) explore 'the use of information obtained from examinations, presumptive and discriminating tests for body fluids, in order to provide assistance to consider propositions at source level.' (p. 54)

As an aside, note also that it has been suggested that when scientists talk about the source of *part* of a DNA profile (i.e. the major component of a profile) they refer to sub-sub-source propositions (Hicks et al., 2016).

The second hierarchical level (level II) is related to activities. This implies that the definition of propositions of interest has to include an action. Activity level propositions could be, for example, 'Mr X kicked the victim' versus 'Mr X did not kick the victim' (i.e. some other man kicked the victim, and Mr X is not involved in the offence), or 'Mr X sat on the car driver's seat' versus 'Mr X never sat on the car driver's seat'. The consequence of this activity, the assault or the sitting on a driver's seat, is an event of contact (between the two people involved in the assault, or the contact between the driver and the seat of the car) and, consequently, a transfer of material (e.g. blood or fibres in this example). The scientist thus needs to consider more detailed information about the case under examination, relative to the transfer, persistence, and recovery of the blood or fibres on the receptor (the victim's pullover or car's seat for example).

The circumstances of the case (e.g. the distance between the victim and the criminal, the strength of the contact and the *modus operandi*, previous and legitimate contacts) need to be known in order to be able to answer relevant questions like 'Is this the sort of trace that would be seen if Mr X were the man who kicked the victim?' or 'Is this the sort of trace that would be seen if Mr X were not the man who kicked the victim (Mr X is unconnected to the assault)? '.

The evaluation of evidence under level I propositions requires little in the way of circumstantial

information. Only I, the background information, is needed. Such information is required in order to define the relevant population for use in the assessment of the rarity of the characteristics of interest (a detailed description of the role of the background information I is given in Section 2.4.4). However, evaluation given activity level (II) propositions is not feasible without also a framework of circumstances (Kokshoorn et al., 2017), notably on how and when traces may have been deposited. The importance of this will clearly appear in the discussion of pre-assessment (see Sections 5.5.1 and 5.5.3). Pre-assessment requires experts to examine the different versions of the case and to verify that all relevant information for the proper assessment of the evidence is available. The main advantage of level II over level I propositions is that the evaluation of evidence under activity (level II) propositions does not strictly depend on the recovered material. It is possible, for example, to assess the fact that no fibres have been recovered and it is clearly important to assess the importance of the absence of material (such absence of material is evidence of interest). A formal development is presented in Sections 5.5.3 and 6.2.6.

Level III, known as the offence level, and sometimes as the crime level, is close to the activity level. It needs elements that qualify the activity as an offence. At level III, the propositions are those of most interest to the jury. Non-scientific information, such as whether or not a crime occurred,

or whether or not an eyewitness is reliable, plays an important role in the inference (on H_p or H_d), and any decision taken subsequently. Factors such as the relevance of the recovered stain, as defined by Stoney (1991a), or the possibility of transfer for innocent reasons play an important role in the assessment of the evidence given level III propositions.

In routine work, forensic scientists predominantly assess scientific evidence given source level propositions. Assessment given activity level propositions requires that an important body of circumstantial information is available to the scientist (see Section 5.5.2). Often this is not the case because of a lack of interaction between the scientists and investigators, or for reasons related to an oversimplification of the case (Biedermann et al., 2016d). There are clear limitations in the use of a sub- or source level evaluation in a criminal investigation compared with an evaluation given activity level propositions. For a discussion on those limitations and on guidelines for biological evidence, see Gill et al. (2018,2020).

Several criminal cases attest to such limitations. See, for example, the discussions reported in the judgement of *R v. Weller* (2010) and the comments on the Amanda Knox case (2015) in Vuille et al. (2013) and Gill (2016). Generally, the lower the level (with offence level being the highest level) at which the evidence is assessed, the more limited

will be the importance of the results in the context of the case as a whole discussed in court. For ease of simplicity, note that even if the value, V, of the evidence is such as to add considerable support to the proposition that the evidence comes from the PoI, this does not help determine whether the evidence had been transferred during the criminal action or for some innocent reason. Consequently, there is often dissatisfaction if the scientist's evaluation is restricted to sub-level I and level I propositions. The ENSFI Guideline (ENFSI, 2015) emphasised this aspect, supporting evaluations at activity level. The guideline specifies:

> *Activity level propositions should be used when expert knowledge is required to consider factors such as transfer mechanisms, persistence and background levels of the material which could have an impact on the understanding of scientific findings relative to the alleged activities. This is particularly important for trace materials such as micro-traces (fibres, glass, gunshot residues, other particles) and small quantities of DNA, drugs or explosives. (p. 11)*

In summary, thus, the available information, the context of the case, and the key issue(s) influence the choice of propositions. Propositions should be amenable to a reasoned assignment of credibility by a judicial body and be useable for rational inference as emphasised in ENFSI (2015). Examples illustrating the various levels of the hierarchy of propositions are given in Table 5.1.

Table 5.1 Examples of the hierarchy of propositions

Level	*Generic*		*Examples*
III	Offence	A	Mr A committed the burglary
			Another person committed the burglary
		B	Mr B raped Ms Y
			Some other man raped Ms Y
		C	Mr C assaulted Mr Z
			Some other man assaulted of Mr Z
II	Activity	A	Mr A is the man who smashed window X
			Another person smashed the window and Mr A was not present when window X was smashed
		B	Mr B had sexual intercourse with Ms Y
			Some other man had sexual intercourse with Ms Y
		C	Mr C is the man who kicked Mr Z in the head
			Some other person kicked Mr Z in the head
I	Source	A	The glass fragments came from window X
			They came from some other broken glass object
		B	The semen came from Mr B
			The semen came from some other man
		C	The blood on Mr C's clothing came from Mr Z
			The blood on Mr C's clothing came from some other (unrelated) person

Source: From Cook et al. (1998a). Reprinted with permissions of Elsevier.

5.3 THE FORMAL DEVELOPMENT OF A LIKELIHOOD RATIO FOR DIFFERENT PROPOSITIONS AND DISCRETE CHARACTERISTICS

5.3.1 Likelihood Ratio with Source Level Propositions

As usual, let E be the evidence, the value of which has to be assessed. Let the two propositions to be compared be denoted H_p and H_d. The likelihood ratio, V, taking into account background information I, is then

$$V = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}.$$

The propositions will be stated explicitly for any particular context. If the court is interested in determining the origin of a given stain or mark, H_p and H_d will explicitly mention that a PoI (or an object) is or is not the source of the stain (mark or signal).

For illustrative purposes, consider a case involving transfer material such as fibres in which a group of fibres has been left at the scene of the crime by the person who committed the crime. Denote by Γ the characteristics of interest. Shortly

after the event, a PoI is arrested by police. A forensic scientist compares the characteristics of the recovered fibres and the characteristics of a group of fibres taken from the PoI's pullover that are also of type Γ. It is of interest, then, to assess the value of this fibre evidence. The two propositions to be considered are:

H_p: the recovered fibres come from the PoI's pullover;

H_d: the recovered fibres come from some other garment.

The scientist's results, denoted by E, may be divided into two parts (E_c, E_r) as follows

E_c: the fibre characteristics, Γ, of the PoI's pullover;

E_r: the characteristics, Γ, of the recovered fibres.

Note that this section assumes the characteristics are discrete. The analysis of continuous measurements on examined materials is presented in Chapter 7. A numerical example is discussed in Chapter 6.

The scientist knows, in addition, from data previously collected that the characteristic Γ occurs in $100\gamma\%$ of some population, Ψ, say. A general formulation of the problem is given here. The value to be attached to E is given by

$$V = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}.$$

This can be simplified.

$$\begin{aligned} V &= \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)} \\ &= \frac{\Pr(E_r, E_c \mid H_p, I)}{\Pr(E_r, E_c \mid H_d, I)} \\ &= \frac{\Pr(E_r \mid E_c, H_p, I)}{\Pr(E_r \mid E_c, H_d, I)} \frac{\Pr(E_c \mid H_p, I)}{\Pr(E_c \mid H_d, I)}. \end{aligned} \quad (5.1)$$

E_c is the evidence that the PoI's pullover has characteristics Γ. An assumption that may be made is that the probability of a person's pullover having characteristics Γ does not depend on whether or not that pullover is the source of the recovered fibres. Thus

$$\Pr(E_c \mid H_p, I) = \Pr(E_c \mid H_d, I)$$

and the likelihood ratio reduces to

$$V = \frac{\Pr(E_r \mid E_c, H_p, I)}{\Pr(E_r \mid E_c, H_d, I)}.$$

If the PoI's pullover is not the source of the recovered fibres (H_d is true), then the evidence (E_r) about the recovered fibres' characteristics is independent of the evidence (E_c) about the characteristics of the PoI's pullover (Aitken and Taroni, 1997). Thus

$$\Pr(E_r \mid E_c, H_d, I) = \Pr(E_r \mid H_d, I)$$

and

$$V = \frac{\Pr(E_r \mid E_c, H_p, I)}{\Pr(E_r \mid H_d, I)}. \quad (5.2)$$

In examples concerning DNA profiles, it will be generally understood that, given a proposition that a stain does not come from a PoI, the stain originated from some person unrelated to the PoI, but from the same sub-population (Balding and Nichols, 1994; Balding and Donnelly, 1995b). This aspect requires a more detailed development of the likelihood ratio because it cannot be assumed that $\Pr(E_r \mid E_c, H_d, I) = \Pr(E_r \mid H_d, I)$. An example is presented in Chapter 6.

The above argument represents a 'suspect-anchored' perspective (Section 5.3.1.3). It is also possible to consider a 'scene-anchored' perspective. A similar argument to that used above shows that

$$V = \frac{\Pr(E_c \mid E_r, H_p, I)}{\Pr(E_c \mid H_d, I)}.$$

This result assumes that $\Pr(E_r \mid H_p, I) = \Pr(E_r \mid H_d, I)$, that is, the features of the recovered material are independent of whether the PoI's pullover was the source of the material at the scene or not, remembering that nothing else is known about the PoI's pullover, in particular its characteristics. This assumption, related there to a PoI's DNA profile, is discussed further in Chapter 6. The assumption that the characteristics of a PoI's garment are independent of whether this garment is the source of the recovered material or not should not be made lightly. Some crime scenes may be likely to transfer materials to an offender's clothing, for example. If

the characteristics of interest relate to such materials and the offender is later identified as a PoI, the presence of such material is not independent of their presence at the crime scene. If the legitimacy of the simplifications is in doubt then the original expression (5.1) is the one which should be used.

The background information, I, may be used to assist in the determination of the relevant population from which the criminal may be supposed to have come. First, consider the numerator $\Pr(E_r \mid E_c, H_p, I)$ of the likelihood ratio in the suspect-anchored perspective. This is the probability the recovered fibres have characteristic Γ given the PoI's pullover is the source of the recovered fibres *and* the PoI's pullover has the characteristic Γ *and* all other information. This probability is 1 since if the PoI's pullover is the source of the recovered material and has the characteristic Γ then the recovered material is of characteristic Γ, assuming as before that all innocent sources of the material have been eliminated and that the characteristics do not change under the prevailing conditions. It is also assumed that the forensic laboratory is able to detect the characteristic Γ on recovered materials every time material having characteristic Γ is collected at crime scenes.

Note that – as explained in Section 6.1.8.4 – observations and measurements are subject to uncertainty and it is important to capture and represent this uncertainty explicitly in the probabilistic model adopted. So, consider E_r as the *reported* observation made by the forensic

scientist and extend further the conditional probability $\Pr(E_r \mid E_c, H_p, I)$ by considering the *true*, but unknown, relationship between control and recovered material (call this event M). The numerator of the likelihood ratio becomes

$$\Pr(E_r \mid E_c, M, H_p, I) \Pr(M \mid E_c, H_p, I) + \Pr(E_r \mid E_c, \bar{M}, H_p, I) \Pr(\bar{M} \mid E_c, H_p, I).$$

Assume that $\Pr(M \mid E_c, H_p, I) = 1$, that is there is no variation in the material of interest. It is also assumed that the PoI's pullover is the source of the recovered material and if the pullover has characteristics Γ, it is (theoretically) certain that the recovered material will match the control material. Thus $\Pr(\bar{M} \mid E_c, H_p, I)$ equals 0. The numerator of the likelihood ratio then equals $\Pr(E_r \mid E_c, M, H_p, I)$, that is, the probability that the forensic scientist will report a correspondence between (E_r and E_c) when a common source is assumed. If it assumed that this probability equals 1, then the numerator becomes 1, as previously specified. A value of 1 for the numerator assumes a correspondence between recovered and control material if there is a common source and an inspection process which has no false negatives. A scenario involving a false positive probability is described in Section 6.1.8.4.

Now consider the denominator, $\Pr(E_r \mid H_d, I)$.

Here the proposition H_d is assumed to be true; i.e. the PoI's pullover is not the source of the recovered material at the crime scene. I is also assumed known. Together I and H_d define the

relevant population (see Section 6.1.3). Several different scenarios are discussed.

5.3.1.1 General Population

For the sake of illustration, consider DNA evidence. Suppose, initially, that I provides no information about the criminal that will affect the probability of his DNA profile being of a particular genotype. For example, I may include eyewitness evidence that the criminal was a tall, young male. However, a DNA profile is independent of all three of these qualities, so I gives no information affecting the probability that the DNA profile is of a particular type.

The alternative proposition assumes that the PoI is not the source of the recovered stain, and hence is not the person who was at the crime scene. The relevant population (Section 6.1.3) is deemed to be Ψ. The donor of the material is an unknown member of Ψ. Evidence E_r is to the effect that the crime stain is of profile Γ. Thus an unknown member of Ψ is Γ. The probability of this is the probability that a person drawn 'at random' (see Section 1.3.2 for a comment on randomness) from Ψ has profile Γ, denote this probability γ. Thus

$$\Pr(E_r \mid H_d, I) = \gamma.$$

Note that this expression assumes no DNA typing error. The laboratory is considered error-free. If this assumption is relaxed the denominator will also consider the false positive probability. A formal development is presented in Section 6.1.8.4.

The likelihood ratio V is then

$$V = \frac{\Pr(E_r \mid E_c, H_p, I)}{\Pr(E_r \mid H_d, I)} = \frac{1}{\gamma}. \qquad (5.3)$$

This value, $1/\gamma$, is the value of the evidence of the profile of the bloodstain when the donor is a member of Ψ. Note that this DNA example considers (unrealistically) E_r and E_c as independent; this is explained Section 6.1.7 for further comment.

5.3.1.2 Particular Population

Suppose now that I does provide information about the donor, relevant to the genotypic occurrence, and the relevant population is now Ψ_0, a subset of Ψ. For example, as mentioned earlier, I may include a eyewitness description of the criminal as Chinese (consider this eyewitness evidence as completely reliable). Suppose the genotype occurrence of Γ amongst Chinese is 100 β%. Then $\Pr(E_r \mid H_d, I) = \beta$ and the likelihood ratio is

$$V = \frac{1}{\beta}.$$

5.3.1.3 Scene- and Suspect-Anchored Perspectives

As mentioned in Section 1.4, it is a feature of likelihood ratios that it is not necessary to distinguish between the scene-anchored and the suspect-anchored perspective; this is explained in the following text and Stoney (1991a).

Consider the following example. A bloodstain is found at the scene of the crime. It is of genotype Γ for a hypothetical locus. All innocent sources of the stain have been eliminated, the knowledge of which may be recorded as relevant background information I. Note here that if it is thought unreasonable to be able to eliminate all innocent sources of the stain then consider, analogously, an example of a rape case in which a semen stain replaces the bloodstain. There will normally be other information which a jury, for example, will have to consider. However, in this context, I is restricted to information considered relevant in the sense that the evidence for which probabilities are of interest is dependent on I (see Section 2.4.4). A PoI has been identified. They are also of genotype Γ. Blood of genotype Γ is not common amongst the general population to which the donor is thought to belong, being found in only 4% of this population. However, the PoI is discovered to be of Ruritanian ethnicity and blood of genotype Γ at a given locus is found in 70% of Ruritanians. How, if at all, should knowledge of the PoI's ethnicity be taken into account? In forensic science it is not uncommon to read that it is appropriate to use reference data on the population from which the PoI originated. This argument is still one of the most persistent fallacies in the DNA debate as mentioned by Weir (1992). The following development will clarify that information on the ethnicity of the PoI does not affect the value of the likelihood ratio. More

Table 5.2 Frequencies of Ruritanians and those of genotype Γ for a given locus in a hypothetical population

	Ruritanians	Others	Total
Genotype Γ	700	100	800
Others	300	18 900	19 200
Total	1 000	19 000	20 000

on the notion of relevant population is presented in Section 6.1.3.

There is assumed at present to be no other evidence, such as eyewitness evidence, to provide information about the ethnic group of the donor of the recovered stain.

Consider Table 5.2, which gives frequencies of Γ in the Ruritanian and other populations. Notice that 800/20 000 (4%) of the general population are of genotype Γ whereas 700/1 000 (70%) of Ruritanians are of genotype Γ, satisfying the description above.

The relevance of the Ruritanian ethnicity of the PoI can be determined by evaluating the likelihood ratio. The evidence E may be partitioned into three parts:

- E_{ru}: the racial grouping (Ruritanian) of the PoI;
- E_c: the genotype (Γ) of the PoI;
- E_r: the genotype (Γ) of the crime stain.

The likelihood ratio is then

$$V = \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} = \frac{\Pr(E_{\text{ru}}, E_c, E_r \mid H_p)}{\Pr(E_{\text{ru}}, E_c, E_r \mid H_d)}$$

where the two propositions of interest are

H_p: the PoI is the source of the recovered material;

H_d: an unknown person is the source of the stain.

Note that in this simplified example it is assumed that there is no genetic relationship between the PoI and the unknown person so that the genetic profiles can be considered independent. Although not explicitly mentioned in the equations, it should be remembered that all the probabilities under discussion here are conditional on I. Notice also that the value of evidence given source level propositions may be thought to imply guilt but the inference of guilt from source is not one which the forensic scientist should make. Rather it is for the jury to make this inference, bearing in mind all other evidence presented at the trial.

Scene-Anchored Perspective The scene-anchored perspective is one in which the evidence related to the PoI (E_{ru}, E_c) is conditioned on the scene evidence E_r. The odds form of Bayes' theorem (2.7)

thus gives

$$V = \frac{\Pr(E_{\text{ru}}, E_c, E_r \mid H_p)}{\Pr(E_{\text{ru}}, E_c, E_r \mid H_d)} = \frac{\Pr(E_{\text{ru}}, E_c \mid E_r, H_p)\Pr(E_r \mid H_p)}{\Pr(E_{\text{ru}}, E_c \mid E_r, H_d)\Pr(E_r \mid H_d)}.$$

Consider, first, the ratio $\Pr(E_r \mid H_p)/\Pr(E_r \mid H_d)$. If no more is assumed of the PoI other than that they were the source of the recovered stain (H_p, the numerator) or that they were not the source of the stain (H_d, the denominator), then the probability of the recovered material being of type Γ is the same whether they were or were not the source of the recovered material, thus $\Pr(E_r \mid H_p) = \Pr(E_r \mid H_d)$. The likelihood ratio reduces to

$$V = \frac{\Pr(E_{\text{ru}}, E_c \mid E_r, H_p)}{\Pr(E_{\text{ru}}, E_c \mid E_r, H_d)}.$$

This may be written as

$$\frac{\Pr(E_{\text{ru}} \mid E_r, H_p)\Pr(E_c \mid E_{\text{ru}}, E_r, H_p)}{\Pr(E_{\text{ru}} \mid E_r, H_d)\Pr(E_c \mid E_{\text{ru}}, E_r, H_d)}.$$

If the PoI was the source of the crime stain (H_p) and if the crime stain is of genotype Γ (E_r), then – assigning a probability with a relative frequency (see Section 1.7.7) – the probability they are Ruritanian is $7/8$, the proportion of Ruritanians amongst those of genotype Γ. Thus $\Pr(E_{\text{ru}} \mid E_r, H_p) = 7/8$.

If the PoI was the source of the recovered material (H_p) and if the crime stain is of genotype Γ (E_r), then the probability the PoI's genotype is Γ (E_c) is

1, independent of his ethnicity (E_{ru}). Thus $\Pr(E_c \mid E_{\text{ru}}, E_r, H_p) = 1$.

If the PoI was not the source of the recovered material (H_d), the blood profile (E_r) of the crime stain gives no information about their ethnicity (E_{ru}). Thus $\Pr(E_{\text{ru}} \mid E_r, H_d) = \Pr(E_{\text{ru}} \mid H_d) = 1/20$, the proportion of Ruritanians in the general population.

Similarly, if the PoI was not the source of the recovered material, knowledge of the features of the crime stain gives no information about the characteristics of the PoI's blood. This, again, assumes that there is no sub-population effect that generates genotyping dependence. Thus $\Pr(E_c \mid E_{\text{ru}}, E_r, H_d) = \Pr(E_c, \mid E_{\text{ru}}, H_d)$ and this is assumed to be the proportion of Ruritanians that are of genotype Γ or, alternatively, the probability that a Ruritanian, selected at random from the population of Ruritanians, is of blood profile Γ. This probability is $7/10$. Then

$$V = \frac{\left(\frac{7}{8}\right) \times 1}{\left(\frac{1}{20}\right) \times \left(\frac{7}{10}\right)} = \frac{7/8}{7/200} = \frac{200}{8}.$$

This is the reciprocal of the proportion of people of genotype Γ in the general population. The ethnicity of the PoI is not relevant. The general proof of this result is given in Evett (1984).

Suspect-Anchored Perspective The suspect-anchored perspective is one in which the scene evidence (E_r) is conditioned on the PoI evidence (E_{ru}, E_c).

From the odds form of Bayes' theorem (2.7)

$$V = \frac{\Pr(E_{\text{ru}}, E_c, E_r \mid H_p)}{\Pr(E_{\text{ru}}, E_c, E_r \mid H_d)} = \frac{\Pr(E_r \mid E_{\text{ru}}, E_c, H_p)\Pr(E_{\text{ru}}, E_c \mid H_p)}{\Pr(E_r \mid E_{\text{ru}}, E_c, H_d)\Pr(E_{\text{ru}}, E_c \mid H_d)}.$$

Consider the ratio $\Pr(E_{\text{ru}}, E_c \mid H_p)/\Pr(E_{\text{ru}}, E_c \mid H_d)$. Assume there is no particular predisposition towards (or away from) criminality amongst Ruritanians (E_r) or those of genotype Γ (E_s). Then, $\Pr(E_{\text{ru}}, E_c \mid H_p) = \Pr(E_{\text{ru}}, E_c \mid H_d)$ and

$$V = \frac{\Pr(E_r \mid E_{\text{ru}}, E_c, H_p,)}{\Pr(E_r \mid E_{\text{ru}}, E_c, H_d)}.$$

The numerator $\Pr(E_r \mid E_{\text{ru}}, E_c, H_p)$ of this ratio equals 1. The PoI is assumed to have been the source of the recovered material (H_p), to be Ruritanian (E_{ru}) and of genotype Γ (E_c). Thus $\Pr(E_r \mid E_{\text{ru}}, E_c, H_p) = 1$.

If the PoI is assumed not to be the source of the recovered material (H_d), the information that they are Ruritanian and of genotype Γ is not of relevance for determining the probability that the crime stain is of type Γ. Thus, the probability in the denominator can be assigned, using the available data, as $800/20\,000\,(8/200)$, which corresponds to the proportion in the general population that has type Γ. Then

$$\Pr(E_r \mid H_d, E_{\text{ru}}, E_c) = 8/200,$$
$$V = 200/8.$$

As before, the ethnicity of the PoI is not relevant. Also, the scene- and suspect-anchored perspectives provide the same result.

5.3.1.4 Some Remarks on Scene- and Suspect-Anchored Perspectives

Suppose now that all innocent sources of the stain have not been eliminated. Consider the scene-anchored perspective. Assume that the PoI was the source of the recovered material (H_p) and that the stain at the crime scene (though not necessarily the 'crime' stain, in that it may not have been left by the criminal) is of group Γ (E_r). No information is contained in this evidence about the ethnicity of the suspect. Thus

$$\Pr(E_{\text{ru}} \mid E_r, H_p) = 1\,000/20\,000 = 1/20.$$

The stain at the crime scene may not have come from the criminal. (See Section 5.3.3.1 for a more detailed discussion of this idea, known as *relevance* (Stoney, 1991a,1994).) Thus, the probability that the genotype of the PoI is Γ, given they are Ruritanian (E_{ru}) and that the PoI was the source of the recovered material (H_p), is assumed to be numerically just the proportion of Γ amongst Ruritanians, which is 700/1 000 (7/10). Thus

$$\Pr(E_c \mid E_{\text{ru}}, E_r, H_p) = 7/10.$$

The probabilities in the denominator have the same values as before, namely, $\Pr(E_{\text{ru}} \mid E_r, H_d) = 1/20$ and $\Pr(E_c \mid E_{\text{ru}}, E_r, H_d) = 7/10$. Hence

$V = 1$. In this case the recovered findings have no probative value. A similar line of argument derives the same result for the suspect-anchored perspective.

Consider now eyewitness evidence that states that the source of the stain is Ruritanian. This eyewitness evidence is assumed to be completely reliable; how to account for evidence that is less than completely reliable is not discussed here (refer to Taroni et al. (2014a) for a discussion of so-called *soft evidence* and to Jeffrey (1983) for a theoretical development. Examples are presented in Corradi et al. (2013) and Garbolino (2014).).

Note, however, that in such a case there are two conditional probabilities to be considered. Let T be an event and let W_T be an eyewitness report of T. Then it is necessary to consider both $\Pr(T \mid W_T)$, the probability the event happened given the eyewitness said it did, and $\Pr(W_T \mid T)$, the probability the eyewitness said the event happened given it did. The purpose of including eyewitness evidence here is to illustrate the effect of restricting the population of potential donors to a particular subgroup of a more general population.

Suppose the relevant background information, I, relates to the ethnicity of the source of the stain. The evidence E is now only of two parts

- E_c: the genotype Γ of the PoI;
- E_r: the genotype Γ of the crime stain.

Evidence E_{ru} of the ethnicity of the PoI has been subsumed into I and is now evidence of the ethnicity of the source.

For the scene-anchored perspective, with I assumed implicitly,

$$V = \frac{\Pr(E_c \mid E_r, H_p)}{\Pr(E_c \mid E_r, H_d)}.$$

The numerator is $\Pr(E_c \mid E_r, H_p) = 1$ since, if it is assumed that the PoI is the donor of the recovered material and the blood profile of the crime stain is Γ, the blood group of the PoI is certain to be Γ also. The denominator is $\Pr(E_c \mid E_r, H_d) = \Pr(E_c \mid H_d) = 7/10$, the proportion of the population of Ruritanians who are of type Γ. Hence, $\mathrm{V} = 10/7$. The eyewitness evidence is such as to ensure that the ethnicity of the PoI (Ruritanian) is relevant.

For the suspect-anchored perspective, again with I assumed implicitly,

$$V = \frac{\Pr(E_r \mid E_c, H_p)}{\Pr(E_r \mid E_c, H_d)}.$$

If the PoI was the source of the recovered material and is of genotype Γ, then the crime stain is certain to be of genotype Γ. Thus, the numerator is $\Pr(E_r \mid E_c, H_p) = 1$. The denominator is $\Pr(E_r \mid E_c, H_d) = \Pr(E_r \mid H_d) = 7/10$ since I includes the information that the source of the stain is Ruritanian. The proportion of the Ruritanian population that is of genotype Γ is $7/10$.

The scene- and suspect-anchored perspectives give the same result.

5.3.1.5 Types of Evidence: Recovered, Control, and Background Data

It is largely accepted – as shown in Sections 2.4 and 2.3.1 – that the likelihood ratio assesses the evidence under competing propositions. The scientific findings or the outcomes of laboratory analyses characterise what has been called evidence. An alternative formulation for the likelihood ratio is the following:

$$V = \frac{\Pr(E_r, E_c, D \mid H_p, I)}{\Pr(E_r, E_c, D \mid H_{d,I})}$$

where D is the background knowledge derived from a population database of, for example, possible donors of the stain or of a mark and is considered part of the evidence. The role of the database is to enable estimation of the population proportion, say, θ, of the feature of interest in the case at hand. So, in a case involving DNA evidence, a given genotype, say, Γ, characterises the recovered stain. The PoI has the Γ profile, and the scientist may be interested in the proportion of people with a Γ profile in a given relevant population.

The first component of the evidence is therefore the recovered and control data that are related to the case under investigation, the second component is information – unrelated to the case – which

is available to the scientist to assess a parameter, θ that is of relevance in the case.

The database information is used by scientists to help evaluate the denominator of the likelihood ratio; see Example 2.7 of Section 2.4.5. Here, this information is directly expressed as relevant data for both the numerator and the denominator.

Imagine a criminal case involving shoemarks recovered on a crime scene. Imagine also that a PoI wears a pair of shoes producing indistinguishable marks if compared with prints recovered at the scene. The marks can be classified as of type T, say. Consider the general situation where there are just two types of marks, those offering an indistinguishable image (T) with probability θ, and those offering a distinguishable image ($\bar{T}$) with probability $(1-\theta)$. Consider also a police shoeprints database of N prints where n are of type T.

The available data are, therefore, shoemarks of type T on the crime scene (E_r), PoI's shoeprints of type T (E_c), and a given number, n, of shoes in the database that are of type T. Independently of the fact that the parameter θ is taken as a fixed but unknown number or as a random variable (see Section 4.2), the likelihood ratio can be viewed as the ratio of the two likelihoods, where, on the numerator the prosecution asserted that there are $n+1$ distinct shoes of type T (the n shoes in the database and the crime shoe which is the same as the PoI shoe) and, on the contrary, the defence (under proposition H_d) supported $n+2$

observations of type T (the n shoes in the database, the crime shoe and the PoI shoe which under H_d is different from the crime shoe). Both parties recognise that $E_r = E_c$ and that exactly n shoes in the database are indistinguishable from E_r and E_c. The available data are conditioned on the propositions and on the population proportion θ. The likelihood ratio can be expressed as

$$\begin{aligned} V &= \frac{\Pr(E_r = T, E_c = T, n, N-n \mid H_p, \theta)}{\Pr(E_r = T, E_c = T, n, N-n \mid H_d, \theta)} \\ &= \frac{\binom{N}{n}}{\binom{N}{n}} \frac{\theta^{n+1}(1-\theta)^{N-1-n+1}}{\theta^{n+2}(1-\theta)^{N-2-n+2}} \\ &= \frac{\binom{N}{n}}{\binom{N}{n}} \frac{\theta^{n+1}(1-\theta)^{N-n}}{\theta^{n+2}(1-\theta)^{N-n}} = \frac{1}{\theta}. \end{aligned}$$

The likelihood ratio simplifies to $1/\theta$, the result presented in Section 5.3.1.1. Other examples of such a development can be found in Ommen et al. (2016), Dawid (2017), and Cereda (2017).

A decision has to be made as to what value of θ to use in the above expression. A Bayesian perspective is presented in Dawid (2017) where the uncertainty on θ is modelled by a probability distribution (see Section 7.2.2 for further details).

This approach considers the background information derived from a population database as

part of the evidence a scientist wishes to evaluate instead of a conditioning event. Note, however, that the database is general information unrelated to a particular crime. The evidence to be evaluated is that associated in particular with the crime under consideration.

5.3.2 Likelihood Ratio with Activity Level Propositions

The transfer, persistence, and recovery of material collected on a receptor (e.g. a person's garments, crime scene, etc.) and the presence by chance of such material on the receptor represent fundamental factors for the evaluation of evidence given *activity level* propositions (see Section 5.2 and Evett (1984), Cook et al. (1993), and Champod and Taroni (2017)). Various technical information that the scientist collects during the analysis, such as (i) the quantity of the recovered material (e.g. the number of recovered fibres), (ii) the materials involved (the material composing the receptor and the potential source), and (iii) the intensity of the posited activity under consideration, are essential for the numerical assignment of the various factors that the likelihood ratio takes into account. Procedures for their quantification have been published, for example, by Chabli (2001) and more recently by Grieve et al. (2017) and Roux

and Wiggins (2017) for scenarios involving fibres evidence.

The influence of such relevant factors on the likelihood ratio is easily shown by the range of the values that can be obtained in different versions of a case (see Chapter 6). Likelihood ratio values can vary over a wide range, from support for the prosecutor's proposition to support for that of the defence. Various versions of cases in which recovered material is assessed given activity level propositions are presented and discussed in forensic literature (e.g. Champod and Taroni, 2017). In fibre transfer cases, examples of propositions of interest could be

H_p: the PoI sat on the driver's seat of the stolen car;

H_{d1}: the PoI never sat on the driver's seat of the stolen car;

H_{d2}: the PoI sat on the seat one week ago for legitimate reasons.

In another context, the propositions could be

H_p: the victim sat on the passenger's seat of the PoI's car;

H_d: the victim has never sat on the passenger's seat of the PoI's car.

With such propositions it can be shown that probabilities for aspects such as the background presence of trace material of interest impact on

the evaluation of the evidence. In particular, it can be shown that the value of the likelihood ratio with activity level propositions simplifies, under certain assumptions, to $1/\theta$ where θ is the proportion of the recovered material in some relevant population.

The following sections present examples to illustrate distinct formal developments associated with different versions of a given case: transfer material left by the offender, Section 5.3.2.1, transfer material not left by the offender, Section 5.3.2.2, and material transferred innocently, Section 5.3.2.3. An extension considering the uncertainty about the source is also introduced, Section 5.3.2.4.

5.3.2.1 Transfer Material Left by an Offender

Imagine a case in which a stolen car is used in a robbery on the day of its theft. One hour after the robbery, the car is abandoned. That night the stolen vehicle is found by the police. On the polyester driver's seats (lower and upper back), a number $n = 170$ of extraneous textile fibres are collected. The day following the robbery, a PoI is apprehended. Their red woollen pullover is seized and submitted to the laboratory.

Following notation introduced in Section 2.4.1, the evidence Ev is the material from the car's seat M_r (where r denotes *recovered*) and from the PoI's pullover M_c (where c denotes *control*) and the characteristics E_r and E_c of these materials. These

characteristics will be denoted y and x, short for E_r and E_c, respectively. The forensic evidence is then described as:

y the group of $n = 170$ red woollen fibres, described by a set y of extrinsic (physical attributes such as quantity and position) and intrinsic characteristics (chemical or physical descriptors such as analytical results);

x the red woollen PoI's pullover generates known fibres described by a set x of intrinsic characteristics.

For a discussion on extrinsic and intrinsic characteristics, see Section 6.2. Note that the evaluation of transfer material given activity level propositions, discussed here, assumes a direct source relationship (i.e. the PoI wore the garment of interest). Situations with uncertainty about the true source are presented in Section 5.3.2.4.

The likelihood ratio is expressed as follows:

$$V = \frac{\Pr(y, x \mid H_p, I)}{\Pr(y, x \mid H_d, I)}$$

where

H_p: The PoI sat on the driver's seat of the stolen car;

H_d: The PoI has never sat on the driver's seat of the stolen car.

The hypothesis H_d implies that an unknown person sat on the driver's seat of the stolen car.

This point is important in the assessment of the probabilities of transfer as will be seen later. The previous equation can be expanded using the third law of probability (1.8)

$$V = \frac{\Pr(y, x \mid H_p, I)}{\Pr(y, x \mid H_d, I)} = \frac{\Pr(y \mid x, H_p, I)}{\Pr(y \mid x, H_d, I)} \times \frac{\Pr(x \mid H_p, I)}{\Pr(x \mid H_d, I)}. \tag{5.4}$$

It is reasonable to assume that the probability of the characteristics of the PoI's pullover, x, do not depend on whether or not the PoI sat on the driver's seat of the stolen car. Thus, the second ratio of the right-hand side of (5.4) equals 1 and the likelihood ratio is reduced to

$$V = \frac{\Pr(y \mid x, H_p, I)}{\Pr(y \mid x, H_d, I)}.$$

It is commonly accepted that the denominator of the likelihood ratio is reduced to $\Pr(y \mid H_d, I)$ because it does not depends on knowledge about the characteristics of the control object (here the PoI's pullover). Note that this is different in the case of DNA evidence (see Section 6.1.7) where the fact that a person is known to share the stain's characteristics influences the assignment of the conditional probability called *random (or conditional) match probability*, sometimes also called *conditional genotype probability*. For an extended discussion on this topic, see Buckleton et al. (2016b,e).

The scientist has to assess (i) the probability of the observed characteristics (both intrinsic and

extrinsic) of the recovered fibres, y, given that the PoI sat on the driver's seat of the stolen car and given that his pullover shares the same forensic characteristics as the fibres found on the car seat (the numerator of V), and (ii) the probability of the observed characteristics of the recovered fibres, y, given that the PoI has never sat on the driver's seat of the stolen car (the denominator of V). In order to assess the findings under these two propositions, it is important to note that the scientist is interested in propositions that imply an activity (the act of sitting on a driver's seat) and thus considers the logical consequence of this activity. Imagine a person, the PoI or the offender (who may be the same person), who sat in the driver's seat. This person and their clothes had a physical contact with the seat, so that fibres from the clothes will have been transferred to the seat (as suggested by Locard's exchange principle). For the successful recovery and analysis of these fibres, it is necessary for them to have persisted on the seat and that they will then be successfully recovered by the forensic scientist. The presence of the evidence of the fibres may be explained in one of two ways.

- The recovered group of 170 (n) fibres was transferred, has persisted, and has been successfully recovered from the driver's seat. In this situation, the group of fibres were not present on the driver's seat before the commission of

the crime. Denote this transfer T_n (or T_{170} in this case).

- The recovered group of 170 (n) fibres was not transferred in the commission of the crime, and hence did not persist on the driver's seat. In this situation, the recovered fibres are unconnected with the action under investigation: the group of fibres were on the driver's seat before the commission of the crime. Denote this absence of transfer T_0.

Note that Gill et al. (2020) commented on potential misunderstandings related to the use of the term *probability of transfer*. They noticed that it is a term that has been used to describe two different concepts. The first is when the term refers to the probability of an activity, a primary/secondary transfer, and is thus an example of the prosecutor's fallacy. However, the term 'transfer' is also used by scientists to designate the probability of a given (relative) quantity of [evidential material] being transferred, having persisted and being recovered if the activity took place. To avoid any misunderstandings, they proposed forensic scientists use the term 'probability of recovering material' for the second concept. However, we retain the original wording, the 'probability of transfer'. With respect to the key principles of interpretation mentioned in Section 5.2 and to avoid misunderstandings recall that the probability t refers to the probability of transfer (or of recovering material), the

probability of a given quantity of material being transferred, having persisted and being recovered. It is not the probability of the occurrence of a primary/secondary transfer.

The two explanations for the presence of the evidence of the fibres may be considered as so-called *Association propositions* (see Section 5.3.3.1). There is an assumption here that, in the commission of the crime, no fibres have been transferred, or that all the fibres that have been transferred are from one source. Inclusion of these two association propositions (see Section 5.3.3.2) and omission of the background information I, for simplicity of notation, leads to

$$V = \frac{\begin{array}{r}\Pr(y \mid x, H_p, T_{170}) \Pr(T_{170} \mid x, H_p) \\ + \Pr(y \mid x, H_p, T_0) \Pr(T_0 \mid x, H_p)\end{array}}{\begin{array}{r}\Pr(y \mid H_d, T_{170}) \Pr(T_{170} \mid H_d) \\ + \Pr(y \mid H_d, T_0) \Pr(T_0 \mid H_d)\end{array}}. \quad (5.5)$$

For this likelihood ratio, consideration needs to be given to eight conditional probabilities.

$\Pr(y \mid x, H_p, T_{170})$ represents the probability of observing and recovering a group of 170 red woollen fibres on the car seat, given that the PoI wore a red woollen pullover, that they sat on the driver's seat of the stolen car and that the group of fibres was transferred during the activity, has persisted, and was recovered successfully. If the PoI sat on the driver's seat and the group has been transferred, this means that the group was not there before the activity. This probability is thus

$1 \times b_0$, where b_0 is the probability of the presence by chance of zero groups of fibres.

$\Pr(T_{170} \mid x, H_p)$ represents the probability that a group of 170 red woollen fibres was transferred, has persisted and was recovered successfully from the driver's seat, given that the PoI sat on the driver's seat of the stolen car. This represents the probability, say, t_{170}, that the fibres had been transferred from the PoI's pullover, had remained, and were recovered. This probability depends on physical characteristics of the PoI's pullover (e.g. sheddability of fibres, the garment's fibre structure). It is assumed that the characteristics are those of the control group because the scientist assesses the probability under H_p. Denote $\Pr(T_{170} \mid x, H_p)$ by t_{170}.

$\Pr(y \mid x, H_p, T_0)$ is the probability that a group of 170 red woollen fibres are recovered from the driver's seat, given that the PoI wore a red woollen pullover, that they sat on the driver's seat of the stolen car and that there was no transfer of fibres during the activity, and hence no persistence and successful recovery. If the group has not been transferred in the commission of the crime, this means that it was present on the seat before the activity. Let $b_{1,m} \times \gamma$ represent the probability of the chance occurrence of a single group of size m, a comparable number of fibres, on the driver's seat ($b_{1,m}$), linked to the relevant population proportion γ for the observed characteristics y.

$\Pr(T_0 \mid x, H_p)$ represents the probability that no group of fibres was transferred, persisted, or

recovered successfully from the PoI's pullover to the driver's seat. This probability, t_0, is assigned assuming that the PoI sat on the driver's seat, H_p.

The numerator of the likelihood ratio is then $b_0 t_{170} + b_{1,m} \gamma t_0$.

Next, consider the terms in the denominator of (5.5).

$\Pr(y \mid H_d, T_{170})$ represents the probability of observing a group of 170 red woollen fibres given that the PoI never sat on the driver's seat of the stolen car and that the group of fibres was transferred, persisted, and recovered successfully during the activity. If the PoI never sat on the driver's seat and the group has been transferred, this means the driver's seat did not have this group of fibres before the commission of the crime and the event of the shared characteristics is one of chance. This probability is $b_0 \times \gamma$.

$\Pr(T_{170} \mid H_d)$ represents the probability that a group of 170 red woollen fibres was transferred, persisted and recovered successfully from the driver's seat given that the PoI never sat on the driver's seat of the stolen car. This means that the probability, say, t'_{170}, has to be assigned assuming that the fibres have been transferred from the offender's garment. Knowledge of the features of the PoI's pullover are thus irrelevant for this assessment, because proposition H_d is that the PoI is not the offender. The probability to be assessed thus depends on the physical characteristics of an unknown garment, that is the one worn by the offender. Denote this probability by t'_{170}.

$\Pr(y \mid H_d, T_0)$ is the probability that a group of 170 red woollen fibres is observed on the driver's seat given that the PoI never sat on the driver's seat and that this group of fibres was not transferred, persisted, or recovered successfully during the activity. If the group of fibres was not transferred, it was present on the seat before the commission of the crime. The probability of the chance occurrence of a group of foreign fibres on the driver's seat linked to the relevant population proportion γ for the observed characteristics y is then $b_{1,m} \times \gamma$.

$\Pr(T_0 \mid H_d)$ represents the probability there was no transfer from the offender's garments to the driver's seat, persistence or recovery of a group of fibres. This probability, t'_0, is assigned assuming that the PoI never sat on the driver's seat and, thus, another individual sat in the stolen car, H_d.

The denominator of the likelihood ratio is then $b_0 \gamma t'_{170} + b_{1,m} \gamma t'_0$ and the likelihood ratio (5.5) is

$$V = \frac{b_0 t_{170} + b_{1,m} \gamma t_0}{b_0 \gamma t'_{170} + b_{1,m} \gamma t'_0}. \tag{5.6}$$

The chosen notation assumes $\Pr(T_{170} \mid x, H_p) \neq \Pr(T_{170} \mid H_d)$ and $\Pr(T_0 \mid x, H_p) \neq \Pr(T_0 \mid H_d)$. In practice, the probabilities are assigned on the basis of results of controlled experiments using the garments involved under propositions H_p and H_d, so it is reasonable to assume that the values are different. Multiple and complex variables involved in transfer, persistence, and recovery

phenomena should be taken into account directly for the assignment of t_{170}, t_0, t'_{170}, and t'_0. The aim is to assign values that appropriately reflect the circumstances of the alleged offence. A modelling technique for the assessment of transfer probabilities in cases involving glass fragments has been developed by Curran et al. (2000) (see also Section 6.2.3.2). The technique can also help in the assignation of probabilities in fibre cases as discussed in Champod and Taroni (2017). Siegel (1997), Roux et al. (1999), and Grieve et al. (2017) discussed the use of surveys to inform probability assignments for different case types involving fibres.

Probabilities for the background presence of fibres, $b_{g,\mathbf{m}}$, where g is the number of discrete groups of fibres and $\mathbf{m}$ is the number of fibres in each of the groups, $\{m_1, \ldots, m_g\}$, may be assigned using data obtained in surveys in which groups of extraneous fibres were recovered from surfaces of interest. The probabilities thus derived depend on the types of fibres considered. In fact, the probabilities are influenced by the conditions of transfer and the sheddability of the potential garments involved (Roux and Margot, 1997). Sometimes so-called target fibres studies are performed. These studies enable the assignment of values for $b_{1,m\gamma}$ directly (Palmer and Chinherende, 1996). The probabilities b_0 and $b_{g,\mathbf{m}}$ are considered as probabilities for mutually exclusive parts of the event 'having 0, 1 or more groups of extraneous fibres

which can be distinguished from the garments of the habitual user(s) of the car' (Champod and Taroni, 2017) such that

$$b_0 + \sum_{g=1}^{\infty} b_{g,\mathbf{m}} = 1.$$

Then:

$$b_{1,\mathbf{m}} = 1 - b_0 - \sum_{g=2}^{\infty} b_{g,\mathbf{m}} \leq 1 - b_0.$$

For practical reasons, as discussed in Champod and Taroni (1997), $b_{1,\mathbf{m}}$ is set as the strict complement of b_0, values for $b_{2,\mathbf{m}}$ to $b_{\infty,\mathbf{m}}$ are set equal to 0 and $\mathbf{m}$ is written as m.

In the example of the car described earlier, it is reasonable to assume that, on average, a large number of extraneous fibres are transferred, have persisted and are recovered from the driver's seat of the stolen car. This implies that t_{170} is much greater than t_0. A similar assumption applies for t'_{170} and t'_0, which are assigned through controlled experiments involving woollen garments of the type potentially worn by the offender. The likelihood ratio (5.6) can then be reduced to

$$V = \frac{t_{170}}{\gamma t'_{170}}$$

because t_0 and t'_0 are assumed negligible and hence $b_{1,m}\gamma t_0$ and $b_{1,m}\gamma t'_0$ are negligible in this case.

Under the particular assumption that the transfer characteristics of the PoI's pullover do not differ

from those of the offender's garment, the likelihood ratio may be reduced further to $1/\gamma$. Lists of references reporting values for γ are given in Chabli (2001), Cantrell et al. (2001), and Palmer (2016).

Imagine two small modifications of the example considered so far. First, consider a modification, notably in the quantity of recovered transfer material (i.e. the number of fibres) on the driver's seat. Second, consider transfer probabilities to differ under propositions H_p and H_d. Thus, $\Pr(T_n \mid x, H_p) \neq \Pr(T_n \mid H_d)$ and $\Pr(T_0 \mid x, H_p) \neq \Pr(T_0 \mid H_d)$. In such situations, Champod and Taroni (2017) have discussed an example in which the following has been observed:

- If a group of 10 fibres has been recovered on the driver's seat and if 10 fibres is the average number expected under the proposition of involvement of the PoI and if, on average, a potential offender's garment will transfer 60 fibres, then the likelihood ratio exceeds the reciprocal of the population proportion, $1/\gamma$, without any notable influence of the background probabilities b_0 and $b_{1,m}$. This is because $V = t_{10}/\gamma t'_{10}$ and t_{10} is close to 1 and t'_{10} is very small, close to zero.
- If a group of 10 fibres has been recovered on the driver's seat and if this quantity does not correspond to the average number expected under the proposition of implication of the PoI (i.e. the PoI's pullover transfers on average 50 fibres) and if on average a potential offender's

garment will transfer a comparable number of fibres (i.e. approximately 10 fibres), then the likelihood ratio may be lower than 1 so that the findings support the defence proposition.

Numerical examples are presented in Chapter 6.

5.3.2.2 Transfer Material Not Left by the Offender

A different situation involving fibres is described later, taken from Champod and Taroni (2017). The case is that of a car belonging to a man who is suspected of abducting a woman. The victim was wearing a red woollen pullover. According to the PoI, nobody sat in the passenger's seat of his car at the time of the alleged event. On this seat, a group of extraneous fibres consisting of $n = 170$ red wool fibres has been collected. In their defence the PoI denied that the victim has ever sat on the passenger's seat of his car. This defence implies that there is no offence at all, so the recovered fibres are not related to a recent activity (e.g. the seating of a person in the passenger's seat of the car). Hence, these fibres are on the passenger's seat by chance alone from an activity that was not recent. This represents an important point for the understanding of the development of the likelihood ratio formula (see also the development presented in Section 6.2). In fact, even if the numerator of the likelihood ratio is still the same as in (5.5), there is a change in the denominator. There is no reason to develop $\Pr(y \mid H_d)$ using

the association propositions T_n and T_0 because the fibres are not considered to be the result of transfer, persistence, and recovery following an alleged action. The likelihood ratio thus becomes

$$V = \frac{b_0 t_{170} + b_{1,m}\gamma t_0}{b_{1,m}\gamma}. \qquad (5.7)$$

It can be seen that if b_0 is close to 1 (e.g. when the seats are cleaned regularly) and if, as mentioned before, $b_{1,m} \simeq 1 - b_0$, then $b_{1,m}$, the probability that the recovered group is present on the seat by chance alone, is close to 0 and the likelihood ratio (5.7) is increased with respect to (5.6). A numerical example is given in Section 6.2.3 in the context of DNA evidence.

The likelihood ratio (5.7) is the method of evaluation widely used in cases involving glass fragments (see Section 6.2.2). This is so because it is often argued that the presence of glass fragments on a PoI's pullover, for example, is the result of chance alone. In fact, H_d normally postulates that the PoI has not broken the window, so an alternative (perhaps criminal) action that could account for the presence of the fragments is not given. Equation (5.7) may be rewritten in this context as

$$V = \frac{b_0 t_{170} + b_{1,m}\gamma t_0}{b_{1,m}\gamma} = \frac{b_0 t_{170}}{b_{1,m}\gamma} + t_0.$$

An extension of this kind of case is given by Buckleton and Evett (1989) where the number,

g, of extraneous groups of fibres (i.e. a number of fibres that is compatible with the nature of the activity) is greater than 1. See also Curran et al. (2000) and Curran and Hicks (2009) for an analogous approach for glass fragments.

Recovered materials on the PoI are denoted $y_1, y_2, \ldots, y_g$. Only one group corresponds to the features of the victim's pullover, x_1. The likelihood ratio developed by Buckleton and Evett (1989) shows that it is important not only to focus on the fibres that match, for example, the victim's garments, but also to consider other groups of fibres that may be compatible with the alleged action. There is an analogy to what is presented in Section 6.1.4 on the so-called two-trace problem.

Buckleton and Evett (1989) showed that

$$V = \frac{\begin{pmatrix} \Pr(y_1, y_2, \ldots, y_g \mid x_1, T_n, H_p) \Pr(T_n \mid x_1, H_p) \\ + \Pr(y_1, y_2, \ldots, y_g \mid x_1, T_0, H_p) \Pr(T_0 \mid x_1, H_p) \end{pmatrix}}{\Pr(y_1, y_2, \ldots, y_g \mid H_d)}. \tag{5.8}$$

The likelihood ratio (5.8) can reasonably be reduced to $t_n/\gamma_1 g$. Details of the development are also given in Champod and Taroni (2017).

5.3.2.3 Innocent Transfer of Material

It may be the case that material is transferred innocently. The transfer may be from the victim to the assailant or from the assailant to the victim or crime scene. In Section 5.3.2.1, it has been noticed that – when considering propositions that imply

an activity – the scientist should pay attention to the logical consequence of these activities. If a person sat in the driver's seat of a car, this person (and their clothes) had a physical contact with the seat; the same phenomenon occurs if a person assaulted a victim. It is of interest to incorporate this consideration explicitly in the formal development as shown in Garbolino and Taroni (2002).

Imagine a scenario involving an assault where characteristics describing the recovered material (say, textile fibres), E_r, found on the PoI, are similar to the characteristics of the control material coming from the victim, E_c. Therefore, define the evidence E as (E_r, E_c) and more precisely by a single group of recovered material of size m.

The main propositions of interest are

H_p: The PoI assaulted the victim;

H_d: Some person other than the PoI assaulted the victim.

The presence of the evidence of the fibres may be explained in one of two ways:

T: There was a transfer from the victim;

$\bar{T}$: There was not a transfer from the victim.

Moreover, an extension can be suggested by taking into account the logical contact caused by the action. If the PoI assaulted the victim, they have a physical contact with the victim. Such

a contact may involve a transfer of evidential material. Define

C: The victim has been in contact with the PoI;

$\bar{C}$: The victim has not been in contact with the PoI.

Consider the likelihood ratio $\mathrm{V} = \Pr(E \mid H_p) / \Pr(E \mid H_d)$ (with background information omitted for ease of notation). As previously noted (see Section 5.3.2.1), the numerator of the likelihood ratio is obtained by 'extending the conversation' to proposition T and $\bar{T}$.

$$\Pr(E \mid H_p) = \Pr(E \mid T, H_p) \Pr(T \mid H_p) + \Pr(E \mid \bar{T}, H_p) \Pr(\bar{T} \mid H_p),$$

where $\Pr(E \mid T, H_p) = b_0$, the probability of no group of extraneous material being present (i.e. an absence of trace material). The conditional probability $\Pr(T \mid H_p)$ needs to take into consideration the uncertainty about propositions C and $\bar{C}$, leading to

$$\Pr(T \mid H_p) = \Pr(T \mid C, H_p) \Pr(C \mid H_p) + \Pr(T \mid \bar{C}, H_p) \Pr(\bar{C} \mid H_p).$$

Consideration needs to be given to at least three conditional probabilities. First, $\Pr(T \mid C, H_p)$ represents the probability that a transfer from the victim occurred. Denote this probability by the letter t.

$\Pr(C \mid H_p)$ is the probability the victim has been in contact with the PoI given that the PoI assaulted

the victim. This probability is c. This probability assignment depends largely on the circumstances of the case, in particular the information on how the assault was committed by the offender.

Note that $\Pr(T \mid \bar{C}, H_p) = 0$; there is no possibility for transfer when there was no contact.

$\Pr(E \mid \bar{T}, H_p)$ equals $b_{1,m}\gamma$, that is, the probability of transfer other than from the victim. In other words, such a probability represents the chance occurrence of a single group of size m on the PoI's clothing linked to the relevant population proportion γ for the observed characteristics.

Finally, $\Pr(\bar{T} \mid H_p)$ is obtained in an extension to C, that is,

$$\begin{aligned}\Pr(\bar{T} \mid H_p) = {} & \Pr(\bar{T} \mid C, H_p)\Pr(C \mid H_p) \\ & + \Pr(\bar{T} \mid \bar{C}, H_p)\Pr(\bar{C} \mid H_p),\end{aligned}$$

where $\Pr(\bar{T} \mid C, H_p) = (1 - t)$, $\Pr(C \mid H_p) = c$, $\Pr(\bar{T} \mid \bar{C}, H_p) = 1$ and $\Pr(\bar{C} \mid H_p) = (1 - c)$.

The numerator of the likelihood ratio becomes

$$\Pr(E \mid H_p) = b_0 tc + b_{1,m}\gamma[(1 - t)c + (1 - c)].$$

Consider the denominator $\Pr(E \mid H_d)$ of the likelihood ratio. It is also obtained by taking into account the uncertainty about propositions T and $\bar{T}$ by writing

$$\begin{aligned}\Pr(E \mid H_d) = {} & \Pr(E \mid T, H_d)\Pr(T \mid H_d) \\ & + \Pr(E \mid \bar{T}, H_d)\Pr(\bar{T} \mid H_d),\end{aligned}$$

where $\Pr(E \mid T, H_p) = b_0$ and the conditional probability $\Pr(T \mid H_d)$ is obtained by

$$\Pr(T \mid H_d) = \Pr(T \mid C, H_d) \Pr(C \mid H_d) + \Pr(T \mid \bar{C}, H_d) \Pr(\bar{C} \mid H_d).$$

$\Pr(T \mid C, H_d) = s$ represents the transfer probability given the alternative proposition. Note that the nature of the crime and the position where the trace was found may lead to a different probability assignments for t and s, under H_p and H_d, respectively.

$\Pr(C \mid H_d) = d$ is a probability that takes into account the possibility that the PoI could have been in contact with the victim for reasons other than the assault.

Note also that $\Pr(T \mid \bar{C}, H_d) = 0$. As under proposition H_p, there is no possibility for transfer when there was no contact.

The denominator of the likelihood ratio becomes

$$\Pr(E \mid H_d) = b_0 sd + b_{1,m}\gamma[(1 - s)d + (1 - d)].$$

The likelihood ratio is

$$V = \frac{b_0 tc + b_{1,m}\gamma[(1 - t)c + (1 - c)]}{b_0 sd + b_{1,m}\gamma[(1 - s)d + (1 - d)]}.$$

From this development, one can easily find that, if $c = 1$ (meaning that it is assumed that the victim has been in contact with the PoI given that the PoI assaulted the victim) and $d = 0$ (meaning that it is assumed that the PoI could not have been in

contact with the victim for reasons other than the assault) (or $s = 0$), then the likelihood ratio becomes as expression (5.7):

$$V = \frac{b_0 t + b_{1,m}\gamma(1-t)}{b_{1,m}\gamma} = \frac{b_0 t_n + b_{1,m}\gamma t_0}{b_{1,m}\gamma},$$

where t_n denotes the recovered group of n fibres that was transferred, has persisted and has been successfully recovered on the PoI, and as a consequence, $t_0 = 1 - t_n$.

Consider, for sake of illustration, the evaluation of recovered material under varying case circumstances.

Imagine a scenario in which a group of fibres has been found on the jacket of a PoI who has been arrested by the police because it is suspected that they physically assaulted the victim. The characteristics of these fibres are distinguishable from the fibres of the PoI's own jacket but indistinguishable from those of the clothing of the victim. Consider H_p as 'The PoI assaulted the victim' and H_d as 'The PoI did not assault the victim'. Probabilities t and s relate to transfer under propositions H_p and H_d, respectively, and are conditioned on background information concerning the circumstances of the case under examination. In fact, if the group of fibres has been found on the PoI's jacket, it may be reasonable to assume that they result from the assault, so

that $t > s$ (i.e. $\Pr(T \mid C, H_p) > \Pr(T \mid C, H_d)$. On the other hand, if fibres are found on the lower part of the trousers, then the assessment might change and the assumption $t = s$ could be made. This reflects the view that the occurrence of a transfer for such a group of fibres would be the same, irrespective of the truth or otherwise of the main propositions H_p and H_d. Notice also that, given the circumstances of the case, it could be reasonable to assume $t = 1$.

Another possibility may be to consider $c = 1$ and $t = 1$ (i.e. no uncertainty about the occurrence of contact and transfer given H_p), and to suppose that the defence strategy is to assume that the PoI was at the scene of the crime for reasons unconnected to the crime and that they had contact with the victim (e.g. they helped the victim after the assault whilst waiting for the police to arrive). The latter assumption leads to $\Pr(C \mid H_d) = d = 1$. The likelihood ratio then becomes

$$V = \frac{b_0}{b_0 s + (1 - s) b_{1,m} \gamma}.$$

When, in addition, it is assumed that $s = 0$, that is, no transfer from the victim occurred under H_d, the likelihood ratio reduces to

$$V = \frac{b_0}{b_{1,m} \gamma}.$$

If one assumes $s = 1$ that transfer occurred from the victim under H_d, the likelihood ratio reduces to

$$V = \frac{b_0}{b_0} = 1.$$

This example illustrates that the circumstances of a case are a fundamental element of the analysis. The probabilities $c, d, t,$ and s are crucial considerations that have a bearing on the probative value. In addition, the probability of a group of fibres being present on the receptor beforehand, denoted $b_{1,m}$, represents a further relevant consideration.

Consider an alternative scenario involving recovered fibres on the seat of a car that belongs to a man who is suspected of abducting a woman and attempting to rape her. There is a single group of foreign red woollen fibres that have been collected from the passenger seat of the car. The victim was wearing a red woollen pullover. According to the PoI, no one other than his wife ever sits on the passenger seat. In addition, the car seats have been vacuumed recently. The PoI denies that the victim has ever been in contact with the car. In such a case, the main issue of concern is H_p, 'The victim sat on the passenger seat of the PoI's car', and H_d, 'The victim has never sat on the passenger seat of the PoI's car'. Proposition C refers to 'The victim has been in contact with the seat' and alternatively, $\bar{C}$, 'The victim has never been in contact with the seat'. Here, it appears reasonable to assume that $\Pr(C \mid H_p) = c = 1$

and $\Pr(C \mid H_d) = d = 0$. The numerator of the likelihood ratio is $b_0 t + (1 - t) b_{1,m} \gamma$. Given H_d, the transfer probability $\Pr(T \mid C, H_d) = s = 0$, so the denominator of the likelihood ratio becomes $b_{1,m} \gamma$. Thus, the likelihood ratio is

$$V = \frac{b_0 t + (1 - t) b_{1,m} \gamma}{b_{1,m} \gamma}.$$

Consider $t = 1$ and the fact that, as mentioned earlier, no one other than the wife ever sits on the passenger seat and also that the car seats have been vacuumed recently. The probability b_0 should therefore be considered as being close to 1. Then, the likelihood ratio becomes approximately $1/b_{1,m} \gamma$.

5.3.2.4 Uncertainty About the True Source

The general approach for the development of a likelihood ratio for the evaluation of evidence given activity level propositions can be applied to various categories of findings (e.g. DNA, fibres, shoemarks). Notice, however, that fibre evidence and shoemarks differ from DNA in the sense that they are not 'intrinsic' to a given individual; in fact, a given individual has, as far as most of the common typing techniques in forensic science are concerned, one and only one DNA profile (leaving aside biological anomalies and other special cases) and it cannot be deliberately modified. Most people, however, almost certainly, have more

than one pullover or more than one pair of shoes. Thus, with such items, it is necessary to make assumptions regarding the relationship between a particular pullover or a given pair of shoes and a particular PoI such as when the shoes or clothes were worn. In the examples presented earlier in this chapter, it was tacitly assumed that there was no uncertainty about the assumed known source. In what follows this assumption is relaxed.

To continue the discussion on transfer evidence, consider again a case involving textile fibres. A potential relationship pertains, in the first place, between the PoI's pullover and fibres recovered on a crime scene. This does not necessarily include a relationship between this PoI and the recovered fibres, and an extended likelihood ratio development is needed in order to account for the possibility of the PoI being (or not being) a wearer of the pullover (Taroni et al., 2012a). Thus, the problem of interest is that of uncertainty about the item itself actually worn by the PoI if they committed the action of interest. It may not be known if the item worn by the PoI during the alleged facts is in fact the item available (and analysed) as a known source. This uncertainty can be phrased in terms of propositions. Moreover, one can no longer refer to the characteristics of the available item of clothing as a control and no longer assume it to be the clothing which was worn by the PoI in the event that they truly are the criminal. A distinction should be made between the item available as a posited known source and

the actual (yet unobserved) source worn by the PoI (under H_p).

Taking into account all these uncertainties, the likelihood ratio presented in (5.6), in a scenario involving a unique group of extraneous fibres, becomes

$$V = \frac{b_0 t_n \delta + b_{1,m} \gamma [t_0 \delta + t_0''](1 - \delta)]}{b_0 \gamma t_n' + b_{1,m} \gamma t_0'}, \qquad (5.9)$$

where δ refers to $[w + \gamma'(1 - w)]$ and w to the probability that the PoI wore the known source (e.g. the seized pullover) at the moment of the alleged event. If the PoI did not wear the known source at the moment of the crime, it is necessary to consider the probability that the garment actually worn by the PoI, but different from the known source, would still be of the same type (e.g. red wool): γ'.

The probability t_0'' refers to the probability of no transfer from the true source, in the event that this source has characteristics different from those seen on the known source (i.e. the PoI's pullover). The factor t_0'' can reasonably be conceptualised as an average probability of no transfer from all potential sources described as different from the PoI's pullover. The probability t_0'' has to be distinguished, however, from t_0 because, potentially, the probability of no transfer from the garment available as a known source may be different. It is also distinguished from the probability t_0', which is reserved for no transfer from the garment

worn by the true offender under the alternative proposition H_d.

Notice that this extension of the numerator of the likelihood ratio has a potential effect only for cases in which there is uncertainty about whether or not the PoI wore the garment available as a known source. If there is no such uncertainty ($w = 1$), then the numerator of the likelihood ratio becomes, as before, $b_0 t + b_{1,m} \gamma t_0$. A complete formal development, including Bayesian network representations, is available in Taroni et al. (2012a).

5.3.2.5 Cross- (or Two-Way) Transfer of Trace Material

Another category of forensically relevant situations relates to cases in which a direct contact between two persons (or objects) or a person and an object may have occurred. In such cases, a so-called cross- or two-way transfer of trace material may take place. Consider again an example with recovered fibres. The example is simplified for generality.

A stolen vehicle is used in a robbery on the day of its theft. An hour later it is abandoned. The vehicle is found by the police a few hours later. The car owner lives alone and has never lent the vehicle to anyone. The owner wears nothing but cotton. The day following the robbery a PoI is

apprehended, their red woollen pullover and their denim jeans are confiscated.[2]

On the driver's black polyester seat, which has recently been cleaned with a car vacuum cleaner, one group of relevant foreign fibres, different from cotton, is collected. It consists of a large number of red woollen fibres. The evidence E_1 is (y_1, x_1) where y_1 refers to the recovered fibres on the car seat and x_1 refers to known (control) material from the PoI's red woollen pullover. The fibres on the driver's seat are assumed to have been transferred from the clothing of the offender to the seat. Foreign fibre groups are groups of fibres that can be distinguished from fibres from a known source, associated either with the PoI or with an object such as a car.

On the PoI's pullover and denim jeans (together), there are many foreign fibre groups. One consists of twenty extraneous black fibres. They are in agreement (in some sense) with the fibres of the driver's seat. The evidence E_2 is (y_2, x_2) where y_2 refers to the twenty recovered fibres on the PoI's

[2] As previously noticed in Section 5.3.2.4, it is generally necessary to make assumptions regarding the relationship between a particular item of clothing and a particular PoI such as to when the clothing was worn. In what follows, it is assumed the woollen pullover and the denim jeans were worn by the PoI at the time of the incident. As there is no uncertainty about this event, no extension considering uncertainty about the true source of the clothes is required.

clothes and x_2 refers to known material from the driver's seat. The competing propositions (at the activity level) could be

H_p: the PoI sat on the driver's seat of the stolen car;

H_d: the PoI has never sat on the driver's seat of the stolen car, they have nothing to do with the robbery or the car theft.

When two individuals, or an individual and an object, such as a car seat, are in contact, a two-way (reciprocal) transfer of material is usually involved. The two sets of recovered traces then have to be considered as dependent. If a transfer has occurred in one direction and the expert has recovered traces characterising this transfer, then the expert would, in general, expect to find trace material characterising transfer in the other direction (Inman and Rudin, 2001). The presence of material characterising transfer in one direction gives information about the presence of material characterising transfer in the other direction. If H_d holds, that is the PoI has never sat on the driver's seat of the stolen car, then knowledge about material found on the car's seat should not affect one's expectations to find material on the PoI's pullover. A formal analysis will clarify these aspects. Note that I is omitted for ease of notation,

so that

$$\frac{\Pr(H_p \mid E_1, E_2)}{\Pr(H_d \mid E_1, E_2)} = \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} \times \frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)}$$
$$= \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} \times \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)}$$
$$\times \frac{\Pr(H_p)}{\Pr(H_d)}. \qquad (5.10)$$

The value of the evidence is then

$$V = \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} \times \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)}. \qquad (5.11)$$

The second ratio is equal to (5.6)

$$\frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} = \frac{b_0 t_n + b_{1,m}\gamma_1 t_0}{b_0 \gamma_1 t'_n + b_{1,m}\gamma_1 t'_0}, \qquad (5.12)$$

where γ_1 is the population proportion for the characteristics from y_1 in extraneous groups of fibres of similar size found on seats of stolen cars, determined with reference to a background database of fibres. Equation (5.12) reduces to $1/\gamma_1$, if two assumptions can be accepted. The transfer probabilities t_i and $t'_i (i = 0, n)$ refer, as previously mentioned in Section 5.3.2.1, to the probabilities of transfer from the PoI's and the true offender's garments, respectively. By considering $t_n = t'_n$ and $t_0 = t'_0$, (5.12) reduces to

$1/\gamma_1$. If these assumptions cannot be made, then the extended form of the likelihood ratio (5.12) holds. Alternatively,

(1) given the case circumstances, one may consider that background material on the car's seat not attributable to the habitual user would be absent; so, $b_{1,m} = 0$ and $b_0 = 1$;
(2) if the PoI has never sat on the driver's seat of the stolen car (proposition H_d), another individual (different from the owner) sat on it; hence, the transfer characteristics of the unknown garment of that individual, the assumed offender, are of importance.

Focus now on the first ratio on the right-hand side in (5.11) that accounts for a recovered group of twenty fibres (y_2) present on the PoI's clothing. In the numerator of (5.11), it is assumed that this group of fibres is, potentially, the result of transfer whilst the PoI sat on the car's seat. In the denominator, the presence of y_2 is considered as being part of the background presence. If the PoI did not sit on the car's seat, the fibres found on their pullover are there by chance alone as previously mentioned in Section 5.3.2.2. This denominator can be written as $b^*_{1,20}\gamma_2$. The * in this notation is used to distinguish the assignment of a probability for the occurrence of fibres on the pullover from some source other than the car seat from the assignment used for the background presence of fibres on the car seat. The subscript 1, 20 refers to 1 group of 20

fibres. The numerator $\Pr(E_2 \mid E_1, H_p)$ may be written as follows

$$\Pr(E_2 \mid E_1, T_2, H_p) \Pr(T_2 \mid E_1, H_p) + \Pr(E_2 \mid E_1, \bar{T}_2, H_p) \Pr(\bar{T}_2 \mid E_1, H_p), \quad (5.13)$$

with an extension of the conversation using events T_2, the transfer of fibres from the car seat to the PoI's pullover, and $\bar{T}_2$, the transfer of no fibres from the car to the PoI's pullover, respectively.

Given T_2, the observation E_1 of the fibres on the car seat (corresponding to the PoI's pullover) does not influence the conditional probability of E_2. Thus, $\Pr(E_2 \mid E_1, T_2, H_p) = \Pr(E_2 \mid T_2, H_p)$ and $\Pr(E_2 \mid E_1, \bar{T}_2, H_p) = \Pr(E_2 \mid \bar{T}_2, H_p)$. Equation (5.13) becomes

$$b_0^* \Pr(T_2 \mid E_1, H_p) + b_{1,20}^* \gamma_2 \Pr(\bar{T}_2 \mid E_1, H_p).$$

Probabilities b_0^* and $b_{1,20}^*$ represent, respectively, zero background and background of one group of comparable size with compatible analytical features.

$\Pr(T_2 \mid E_1, H_p)$ can also be extended by considering the events of transfer T_1, the transfer of fibres from the PoI's (offender's) pullover to the car seat, and $\bar{T}_1$, the transfer of no fibres from the PoI's (offender's) pullover to the car seat, so that

$$\Pr(T_2 \mid E_1, H_p) = \Pr(T_2 \mid T_1, H_p) \Pr(T_1 \mid E_1, H_p) + \Pr(T_2 \mid \bar{T}_1, H_p) \Pr(\bar{T}_1 \mid E_1, H_p),$$

because it is considered that event of transfer T_2 and E_1 are conditionally independent given T_1.

Denote the probability $\Pr(T_2 \mid T_1, H_p)$ as $u_{n|T_1}$. This is a conditional transfer probability (i.e. conditional on the event T_1 of transfer of a group of n foreign fibres to the car seat). It is highly case dependent because it is influenced by the kind of textile material involved in the transfer.

Probability $\Pr(T_1 \mid E_1, H_p)$ is obtained using Bayes' theorem:

$$\Pr(T_1 \mid E_1, H_p) = \frac{\Pr(E_1 \mid T_1, H_p)\Pr(T_1 \mid H_p)}{\left(\begin{array}{c}\Pr(E_1 \mid T_1, H_p)\Pr(T_1 \mid H_p) \\ +\Pr(E_1 \mid \bar{T}_1, H_p)\Pr(\bar{T}_1 \mid H_p)\end{array}\right)}$$

$$= \frac{b_0 t_n}{b_0 t_n + b_{1,m}\gamma_1 t_0}. \qquad (5.14)$$

For simplicity, consider again the previously supposed extreme situation with values $b_0 = 1$ and $b_{1,m} = 0$ for the background presence of extraneous fibres on the driver's seat. In such a situation, $\Pr(T_1 \mid E_1, H_p) = 1$ and therefore $\Pr(\bar{T}_1 \mid E_1, H_p) = 1 - \Pr(T_1 \mid E_1, H_p) = 0$. This expresses the view that if there were no background fibres on the driver's seat, but fibres corresponding to those on the PoI's pullover were found on the seat, then it is the event of transfer that led to this finding. Thus, $\Pr(T_2 \mid E_1, H_p) = u_{n|T_1}$ and $\Pr(\bar{T}_2 \mid E_1, H_p) = (1 - u_{n|T_1})$ and (5.13) becomes

$$b^*_0 u_{20|T_1} + b^*_{1,20}\gamma_2(1 - u_{20|T_1}).$$

Notice that the probability $u_{20|\bar{T}_1}$ does not appear in this expression due to the extreme values assumed for b_0 and $b_{1,20}$. If these assumptions are relaxed, then the probability $u_{20|\bar{T}_1}$ will appear in the numerator of the likelihood ratio. The denominator $\Pr(\mathrm{E}_2 \mid \mathrm{E}_1, \mathrm{H}_d) = \mathrm{b}^*_{1,20}\gamma_2$. By combining the developments for the numerator and denominator of the likelihood ratios for the two parts of (5.13) the final expression is

$$\frac{b^*_0 u_{20|T_1} + b^*_{1,20}\gamma_2(1 - u_{20|T_1})}{b^*_{1,20}\gamma_2} \times \frac{1}{\gamma_1}. \qquad (5.15)$$

Treatment of the scenario in which there is a two-way transfer of evidence is a particular example of what is known generally as the problem of 'combining items of evidence'. More on this problem is presented in Section 5.6.

5.3.3 Likelihood Ratio with Offence Level Propositions

An extension of the results of Sections 5.3.1 and 5.3.2 to deal with two further issues is considered by Evett (1993a) and refers to the development of a likelihood ratio with offence level propositions. The first issue is that of *relevance* in relation to material. Relevance in this context is defined by Stoney (1991a) and Stoney (1994) as follows. Crime material that is known to come from the offender is said to be relevant in that it is relevant to the consideration of persons of interest as

possible offenders. Here, the notion of relevant material should be distinguished from that of relevant populations. The second issue concerns the recognition that if the material is not relevant to the case then it may have arrived at the scene from a PoI for innocent reasons. In this section, reference to I has, in general, been omitted for ease of notation.

Evett (1993a) considers a situation in which material is found on a crime scene. A crime has been committed by k (≥ 1) offenders. A single bloodstain is found at the crime scene in a position where it may have been left by one of the offenders. A PoI is found and they give a blood sample. The PoI's sample and the crime stain are of the same profile, say, Γ. This profile is shared by the proportion γ of the relevant population from which the k offenders have come.

Suppose that the court is interested in the evaluation of this evidence given the following offence-level propositions:

H_p: the PoI is one of the k offenders;

H_d: the PoI is not one of the k offenders.

Notice the difference between these propositions and those of Section 5.3.1. There, the propositions referred to the PoI (or an object) being, or not being, the source (donor) of the evidential material found at the crime scene. Now, the propositions are stronger, namely, that the PoI is, or is not, one

of the offenders. The value V of the evidence is

$$V = \frac{\Pr(E_r \mid H_p, E_c)}{\Pr(E_r \mid H_d)} \tag{5.16}$$

where E_r is the profile Γ of the crime stain and E_c is the profile Γ of the PoI, sometimes written as $E_r = \Gamma, E_c = \Gamma$, respectively. Note, again, that E_c is considered not to influence the assessment under the alternative proposition H_d.

5.3.3.1 Probabilities of Innocent Acquisition and Relevance

An argument needs to be developed, based on what has been observed – the stain at the crime scene – and the propositions of interest, that the PoI is or is not one of the offenders. The argument is made in two steps.

The first step is the consideration of a proposition according to which the crime stain came from one of the k offenders and an alternative proposition that the crime stain did not come from any of the k offenders. These propositions are known as *association propositions*, or *association hypotheses* (Buckleton, personal communication, cited by (Evett, 1993a)) and introduced in (5.5).

Assume that the crime stain came from one of the k offenders. The second step of the argument is the consideration of a proposition that the crime stain came from the PoI and the alternative

proposition that the crime stain did not come from the PoI. These propositions are known as *intermediate association propositions*.

Development of these propositions requires consideration of other factors. These are *innocent acquisition* and *relevance*. The evaluation of these factors may be done by partitioning the expressions in the numerator and denominator of (5.16). The following subjective (or personal) probabilities are of interest:

- that of innocent acquisition, usually denoted p: this value is a measure of belief that evidence has been acquired in a manner unrelated to the crime (Evett, 1993a);
- that of relevance, usually denoted r (Stoney, 1991a,1994; Evett et al., 1998). In this context it denotes the probability that the stain recovered from the crime scene is connected with the crime and hence has been left by one of the offenders.

5.3.3.2 Association Propositions

Consider the following association propositions:

B : the crime stain came from one of the k offenders;

$\bar{B}$: the crime stain did not come from any of the k offenders.

The value, V, of the evidence may now be written using the law of total probability (Section

1.7.10) as

$$V = \frac{\begin{array}{c}\Pr(E_r \mid H_p, B, E_c)\Pr(B \mid H_p, E_c) \\ + \Pr(E_r \mid H_p, \bar{B}, E_c)\Pr(\bar{B} \mid H_p, E_c)\end{array}}{\begin{array}{c}\Pr(E_r \mid H_d, B)\Pr(B \mid H_d) \\ + \Pr(E_r \mid H_d, \bar{B})\Pr(\bar{B} \mid H_d)\end{array}}.$$

In the absence of information E_r, regarding the profile of the crime stain, knowledge of H_p and of E_c does not affect our belief in the truth or otherwise of B. This is what is meant by *relevance* in this context. Thus

$$\Pr(B \mid H_p, E_c) = \Pr(B \mid H_p) = \Pr(B)$$

and

$$\Pr(\bar{B} \mid H_p, E_c) = \Pr(\bar{B} \mid H_p) = \Pr(\bar{B}).$$

Let $\Pr(B) = r$ and $\Pr(\bar{B}) = (1 - r)$, and call r the *relevance term*. It is the probability that the stain has been left by one of the offenders. The higher the value of r, the more the stain is considered relevant. Thus

$$V = \frac{\Pr(E_r \mid H_p, B, E_c)r + \Pr(E_r \mid H_p, \bar{B}, E_c)(1 - r)}{\Pr(E_r \mid H_d, B, E_c)r + \Pr(E_r \mid H_d, \bar{B}, E_c)(1 - r)}. \tag{5.17}$$

5.3.3.3 Intermediate Association Propositions

In order to determine the component probabilities of (5.17), the following intermediate association propositions are introduced:

A: the crime stain came from the PoI;

$\bar{A}$: the crime stain did not come from the PoI.

Now consider the four conditional probabilities from (5.17).

(a) $\Pr(E_r \mid H_p, B, E_c)$: This is the probability that the crime stain would be of profile Γ if it had been left by one of the offenders (B), the PoI was one of the k offenders (H_p) and the PoI is of profile $\Gamma(\mathrm{E_c})$.

$$\begin{aligned}&\Pr(E_r \mid H_p, B, E_c)\\ &\quad = \Pr(E_r \mid H_p, B, A, E_c)\Pr(A \mid H_p, B, E_c)\\ &\qquad + \Pr(E_r \mid H_p, B, \bar{A}, E_c)\Pr(\bar{A} \mid H_p, B, E_c).\end{aligned}$$

Here $E_r = E_c = \Gamma$ and $\Pr(E_r \mid H_p, B, A, E_c) = 1$. In the absence of E_r, A is independent of E_c and so

$$\Pr(A \mid H_p, B, E_c) = \Pr(A \mid H_p, B) = 1/k,$$

assuming that there is nothing in the background information I to distinguish the PoI, given H_p, from the other $(k-1)$ offenders as far as blood shedding is considered.

In a similar manner, $\Pr(\bar{A} \mid H_p, B, E_c) = (k-1)/k$. Also,

$$\Pr(E_r \mid H_p, B, \bar{A}, E_c) = \Pr(E_r \mid H_p, B, \bar{A}) = \gamma,$$

since if $\bar{A}$ is true, E_r and E_c are independent, and one of the other offenders left the stain

(since B holds). Thus

$$\Pr(E_r \mid H_p, B, E_c) = \{1 + (k-1)\gamma\}/k.$$

(b) $\Pr(E_r \mid H_p, \bar{B}, E_c)$: This is the probability that the crime stain would be of profile Γ if it had been left by an unknown person who was unconnected with the crime (i.e. none of the k offenders). This is the implication of assuming $\bar{B}$ to be true. The population of people who may have left the stain is not necessarily the same as the population from which the criminals are assumed to have come. Thus, let

$$\Pr(E_r \mid H_p, \bar{B}, E_c) = \gamma'.$$

where γ' is the probability of observing profile Γ in a given person from the population of people who may have left the stain. Note that the symbol $'$ is used to indicate that it may not be the same value as γ, which relates to the population from which the criminals have come.

Consider now that the PoI is not one of the k offenders and H_d is true.

(c) $\Pr(E_r \mid H_d, B, E_c) = \Pr(E_r \mid H_d, B) = \gamma$ designates the probability of observing profile Γ by chance on a given person from the population from which the criminals have come. There is no need to partition these probabilities to consider A and $\bar{A}$ as the PoI is assumed not to be one of the offenders and B is that the stain was left by one of the offenders.

(d) Consider now the term

$$\begin{aligned}&\Pr(E_r \mid H_d, \bar{B}, E_c)\\ &\quad = \Pr(E_r \mid H_d, \bar{B}, A, E_c)\Pr(A \mid H_d, \bar{B}, E_c)\\ &\qquad + \Pr(E_r \mid H_d, \bar{B}, \bar{A}, E_c)\Pr(\bar{A} \mid H_d, \bar{B}, E_c).\end{aligned}$$

If A is true, $\Pr(E_r \mid H_d, \bar{B}, A, E_c) = 1$. Also $\Pr(A \mid H_d, \bar{B}, E_c) = \Pr(A \mid H_d, \bar{B})$. This is the probability p of innocent acquisition: the probability that the stain would have been left by the PoI even though the PoI was not one of the k offenders. It is assumed that the propensity to leave a stain is independent of the profile of the person who left the stain. Hence $\Pr(A \mid H_d, \bar{B}) = p$ and $\Pr(\bar{A} \mid H_d, \bar{B}, E_c) = \Pr(\bar{A} \mid H_d, \bar{B}) = 1 - p$. Also $\Pr(E_r \mid H_d, \bar{B}, \bar{A}) = \gamma'$. Thus

$$\Pr(E_r \mid H_d, \bar{B}, E_c) = p + (1-p)\gamma'.$$

Substitution of the aforementioned expressions into (5.17) gives

$$\begin{aligned}V &= \frac{[r\{1+(k-1)\gamma\}/k] + \{\gamma'(1-r)\}}{\gamma r + \{p + (1-p)\gamma'\}(1-r)}\\ &= \frac{r\{1+(k-1)\gamma\} + k\gamma'(1-r)}{k[\gamma r + \{p+(1-p)\gamma'\}(1-r)]},\end{aligned} \tag{5.18}$$

as developed by Evett (1993a). An example of such an evaluation is developed in Chapter 6.3.1.

5.3.3.4 A Note on Evidence Assessment Given Offence Level Propositions

As noted previously (Section 5.3.3.2), association propositions in cases with material found on the crime scene are of the kind 'the crime stain comes (does not come) from the offender' (Evett, 1993a). For cases involving transfer away from the scene (e.g. from victim to offender), the relevance of staining found on a PoI cannot be established analogously by saying, e.g. that 'the stain, found on the PoI, comes (does not come) from the victim'. This is a source level proposition (the victim is stated as the source of the material recovered on the PoI). As such, it is not sufficient to establish an argumentative connection from the observed correspondence between E_c and E_r to the offence level propositions H. In fact, for cases involving transfer away from the scene, it appears reasonable to assume that the relevance of a stain found on a PoI depends on whether or not the particular category of evidence to which the stain belongs (e.g. blood) was actually produced during the course of the offence, depending on what is known through the framework of circumstances. In a victim (or scene) to offender transfer setting, establishment that shedding of material (e.g. by bleeding) is relevant to the case and an assumption that the PoI is the offender do not necessarily imply that the stain found on the PoI comes from the victim.

This implication depends on whether or not there is background staining on the PoI and whether or not transfer from scene to offender occurred.

Relevance in this context could thus be interpreted as a property of the type (or category) of evidence, such as blood, rather than the stain itself. Although an alternative way to relate the stain at hand to the crime could consist of a proposition of the kind 'the non-self stain found on the PoI is present as a result of the crime', this proposition can be thought to be equivalent to the proposition that the stain comes from the victim. The result is a hidden redundancy with respect to a genuine source-level proposition.

A tentative approach to deal with offence level propositions in a case involving potential transfer of material from a victim to a PoI is presented in Biedermann and Taroni (2011), using Bayesian networks to deal with the complexities of formal developments.

5.4 VALIDATION OF BAYESIAN NETWORK STRUCTURES: AN EXAMPLE

Consider, for the purpose of illustration, the case involving a recovered stain at a crime scene and a PoI as a potential offender. The evaluation of evidence given offence-level propositions has been described earlier (see Section 5.3.3). To approach

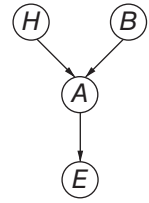

Figure 5.1 Four-node network for the evaluation of evidence given offence level propositions.

this issue probabilistically, it has been shown that a connection is needed between the nodes representing the observation made on the crime stain and the main proposition according to which the PoI is the offender. The connection is made in two steps. The first is the consideration of the proposition that the crime stain came from the offender (i.e. the association proposition). Then, assuming that the crime stain came from the offender, the second step is the consideration of the proposition that the crime stain came from the PoI (i.e. the intermediate association proposition). The Bayesian network (see Section 2.9.1) describing this situation contains four nodes as shown in Figure 5.1.

It is assumed that all four nodes are binary. The two possible values for each of the nodes are as follows:

- H: the PoI is (H_p) or is not the offender (H_d);
- B: the crime stain did (B) or did not come from the offender ($\bar{B}$);
- A: the crime stain did (A) or did not come from the PoI ($\bar{A}$);

- E: the PoI and crime stain have (E) or do not have ($\bar{E}$) the same DNA profile.

Nodes H and B are parent nodes and are independent, i.e. there is no link between these two nodes. Knowing that the stain comes from the offender does not tell one anything about the probability that the PoI is (or is not) the offender. Thus, only one probability needs to be specified for each node, i.e. $\Pr(H_p)$, the probability the PoI is the offender, and $\Pr(B)$, the probability the crime stain came from the offender (i.e. the so-called relevance term). The probabilities $\Pr(H_d)$ and $\Pr(\bar{B})$ are the complements of $\Pr(H_p)$ and $\Pr(B)$, respectively. The outcome of node A is dependent on the nodes H and B. Four probabilities are needed:

- $\Pr(A \mid H_p, B)$: The probability that the crime stain came from the PoI, conditional on the PoI being the offender and the crime stain coming from the offender; here, certainly, the stain came from the PoI, so the probability equals 1.
- $\Pr(A \mid H_p, \bar{B})$: The probability that the crime stain came from the PoI, conditional on the PoI being the offender and the crime stain not coming from the offender; here, certainly, the stain did not come from the PoI, so the probability equals 0.
- $\Pr(A \mid H_d, B)$: The probability that the crime stain came from the PoI, conditional on the PoI not being the offender and the crime stain

coming from the offender; here, certainly, the stain did not come from the PoI, so the probability equals 0.

- $\Pr(A \mid H_d, \bar{B})$: The probability that the crime stain came from the PoI, conditional on the PoI not being the offender and the crime stain not coming from the offender; this is the probability that the stain would have been left by the PoI even though they were innocent of the offence (this probability is denoted p in Section 5.3.3.3).

For the fourth node E two probabilities need to be assigned. The first is $\Pr(E \mid A)$, the probability the PoI and the crime stain have the same profile, given the crime stain came from the PoI. This probability is assigned the value 1. The second is $\Pr(E \mid \bar{A})$, the probability the PoI and the crime stain have the same profile, given the crime stain did not come from the PoI. This is the profile probability γ, assigned using data from the relevant population.

The probability of interest is $\Pr(H_p \mid E)$. The probabilities given earlier are provided. There is then an observation that node E takes the value 'PoI and crime stain have the same DNA profile'. Then

$$\begin{aligned}\Pr(H_p \mid E) &= \Pr(E \mid H_p)\Pr(H_p)/\Pr(E) \\ &= \frac{\Pr(E \mid H_p)\Pr(H_p)}{\Pr(E \mid A)\Pr(A) + \Pr(E \mid \bar{A})\Pr(\bar{A})}\,.\end{aligned}$$

The probability of A can be determined as

$$\Pr(A \mid H_p, B)\Pr(H_p, B) + \Pr(A \mid H_p, \bar{B})\Pr(H_p, \bar{B}) + \Pr(A \mid H_d, B)\Pr(H_d, B) + \Pr(A \mid H_d, \bar{B})\Pr(H_d, \bar{B})$$

which, making use of the independence of H and B, can be rewritten as

$$\begin{aligned}&\Pr(A \mid H_p, B)\Pr(H_p)\Pr(B)\\ &+ \Pr(A \mid H_p, \bar{B})\Pr(H_p)\Pr(\bar{B})\\ &+ \Pr(A \mid H_d, B)\Pr(H_d)\Pr(B)\\ &+ \Pr(A \mid H_d, \bar{B})\Pr(H_d)\Pr(\bar{B}).\end{aligned}$$

Then $\Pr(E \mid H_p)$ can be determined as

$$\begin{aligned}\Pr(E \mid H_p) &= \Pr(E \mid H_p, A)\Pr(A \mid H_p)\\ &\quad + \Pr(E \mid H_p, \bar{A})\Pr(\bar{A} \mid H_p)\\ &= \Pr(E \mid A)\Pr(A \mid H_p)\\ &\quad + \Pr(E \mid \bar{A})\Pr(\bar{A} \mid H_p),\end{aligned}$$

and

$$\begin{aligned}\Pr(A \mid H_p) &= \Pr(A \mid H_p, B)\Pr(B)\\ &\quad + \Pr(A \mid H_p, \bar{B})\Pr(\bar{B})\end{aligned}$$

with $\Pr(\bar{A} \mid H_p) = 1 - \Pr(A \mid H_p)$.

If it is assumed that $\Pr(B) = r$ and $\Pr(A \mid \bar{B}, H_d) = p$, a simplified version of (5.18) (when the number of offenders is reduced to $k = 1$), is obtained:

$$V = \frac{r + \{\gamma'(1 - r)\}}{\gamma r + [p + (1 - p)\gamma'](1 - r)}. \qquad (5.19)$$

Note that r defines the relevance term, p is the probability of innocent acquisition, γ denotes the probability of observing the profile of interest by chance on a given selected person from the population from which the criminal has come, and finally γ' represents the probability of observing the profile of interest in a given selected person from the population of people who may have left the stain (for an innocent reason).

The agreement between the likelihood ratio derived from the Bayesian network and an existing likelihood ratio formula (here (5.18), previously published in Garbolino and Taroni (2002)) can be taken as an indication that the Bayesian network structure and the associated probabilistic assessments are appropriate. Consider also that if it is assumed further that p equals 0 and that the relevance of the crime stain is maximal ($r = 1$), then the likelihood ratio is reduced to its simplest form, $1/\gamma$, a well-known result in the context of evaluation given source level propositions. Garbolino and Taroni (2002) gave further examples of Bayesian network structures that accurately present existing and accepted probabilistic solutions for forensic inference problems.

Section 5.5 introduces the notion of *pre-assessment* and a real-case application. Scientists are encouraged with this approach to consider what propositions they should address before attempting any examination and considering what magnitude of the likelihood ratio

might be expected if the prosecution or the defence proposition were true (Champod and Evett, 2009).

5.5 PRE-ASSESSMENT

5.5.1 Pre-assessment of the Case

The evaluation process should start when the scientist first meets the case. It is at this stage that the scientist thinks about the issues with which they are requested to help and the outcomes that may be expected. The scientist should provide assistance in the definitions of the propositions of interest and think about the value of evidence that is expected (Evett et al., 2000e). However, there is a tendency to consider evaluation of evidence as a final step of casework examination, notably at the time of preparing the formal evaluative report. This is so even if an earlier interest in the process would enable the scientist to make better decisions about the allocation of resources (Jackson et al., 2006). For example, consider a case of assault involving the possible cross-transfer of textile fibres between a victim and assailant. The scientist has to decide whether to look first for potentially transferred fibres on the victim's pullover, or for extraneous fibres potentially present on a PoI's pullover. If traces analytically consistent with the PoI's pullover are found on the victim's pullover, then the expectation of the detection of traces

analytically consistent with the victim's pullover on the PoI's pullover has to be assessed. This includes the possibility of reciprocal transfer. Should the scientist have expectations? How can they be quantified? And, what is the interpretative consequence when those expectations are or are not met, that is, expected material is present or absent? Matters to be considered here include (i) an adequate definition of expectation, (ii) the quantification of the expectations, and (iii) the interpretation of the presence or absence of trace material, in terms of the realisation or otherwise of the expectations.

The scientist requires an adequate appreciation of the circumstances of the case so that a framework may be set up for consideration of the kind of examinations that may be carried out and what may be expected from them (Cook et al., 1998a) and (Champod and Evett, 2009), in order for a logical decision to be made (Taroni et al., 2005).

Such a procedure is known as a *pre-assessment* of the case. It can be justified on a variety of grounds. An essential argument is that the choice of the appropriate level of proposition for the evaluation of scientific evidence is carried out within a framework of circumstances. These circumstances have to be known before any examination can be made in order that meaningful and helpful propositions may be proposed. This is particularly important when a choice needs to be made between activity and source level propositions. As noticed by the

ENFSI Guideline (ENFSI, 2015):

> *Conditioning information is the relevant case information that helps the forensic practitioner recognise the pertinent issues, select the appropriate propositions and carry out the case pre-assessment. It shall always be regarded as provisional and the examiner shall be ready to re-evaluate findings if the conditioning information changes. Examples of relevant information that could change include the nature of the alleged activities, time interval between incident and the collection of traces (and reference items) and the PoI's / victim's account of their activities. (p. 21)*

Moreover, this process provides a basis for a consistency of approach by all scientists who are thereby encouraged to consider carefully factors such as circumstantial information and data that are to be used for the evaluation of evidence, and to declare these considerations in the final report (Jackson and Jones, 2009).

The ENFSI Guideline (ENFSI, 2015) reinforces the importance of the pre-assessment procedure by affirming:

> *Case pre-assessment seeks to specify potential findings prior to performing any analyses or prior to knowing the results, in order to assess the potential value associated with each of these findings, as well as the probability with which these results may be obtained under each of the competing propositions. The purpose is to (i) avoid bias in the evaluations of the findings, and (ii) devise an examination strategy on which a mandating authority or party can – in terms of expected results and associated evidential value – agree (Cook et al., 1998a). To ensure a balanced approach, forensic practitioners should – prior to any examinations – formulate potential outcomes (along*

> *with probabilities for these outcomes) given, in turn, that each of the competing propositions is true. Otherwise an evaluation may be biased. For example, a statement of the kind: 'These observations correspond well to my expectations if the prosecutor's proposition is true' is more trustworthy if the scientist can demonstrate that the respective expectations (including assignments for factors such as transfer and persistence) have been formulated prior to conducting any examinations. (pp. 22–23)*

The scientist should proceed by providing probability assignments for whatever evidence will be found given each proposition of interest. Consider, for example, a case in which a window was smashed and assume that the prosecution and defence propose the following competing versions of the event: 'The PoI is the person who smashed the window', and 'The PoI was not present when the window was smashed'. The examination of the PoI's pullover will reveal a quantity, say, Q, of glass fragments, where Q can be, for example, one of the following states {*no, few, many*} or a list of numerical ranges such as {0, 1–10, 11–100, > 100}. The role of these categories is to maximise the possibility of discrimination between the propositions of interest. Next, the scientist should consider the following two questions.

- The first question focuses on the assessment of a value for the numerator of the likelihood ratio: 'What is the probability of finding a quantity Q of matching glass fragments on the PoI's pullover if the PoI is the man who smashed the window?'

- The second question focuses on the assessment of a value for the denominator of the likelihood ratio: 'What is the probability of finding a quantity Q of glass fragments (with corresponding analytical features) on the PoI's pullover if the PoI was not the person who smashed the window (moreover, the PoI was not present when the window was smashed)?'

This leads the scientist to assess six different probabilities, related to three potential findings, such as {*no, few, many*} (fragments), and two propositions of interest (H_p and H_d), using data on surveys, relevant publications on the matter, or personal expert assessments: the probability of finding {*no, few, many*} matching glass fragments if the PoI is the person who smashed the window, and the probability of finding {*no, few, many*} matching glass fragments if the PoI was not present when the window was smashed.

These probabilities may not be easy to assign because of a lack of information available to the scientist. For example, it will be very difficult to assess transfer probabilities (see Section 6.2.2) if the scientist has no answer to questions like the following.

- Was the window smashed by a person or by a vehicle? The fact that a window was smashed by a person or by a vehicle changes the amount of glass fragments the scientist expects to be transferred.

- How (*modus operandi*) was the window smashed? If it was smashed by a person, then was that person standing close to it? Was a brick thrown through it? Was a hammer used, and if so how many blows were necessary to break the window? Information about the way a window is smashed is important because it provides information on the amount of glass potentially projected. Information of the distance between the person who smashed the window and the window offers relevant information on the amount of glass fragments the scientist will expect to recover.

Where there is little information about the time of the alleged offence and the time at which investigators took possession of any clothing that may be associated with the alleged offence, the lapse of time between the offence and the collection of transfer material cannot precisely be assessed. It is also difficult to assign the probability of persistence of any transferred glass fragments (see Section 6.2.3). Therefore, if the scientist has a limited amount of information about the case under examination, then the pre-assessment has to clearly mention this fact. Consideration only of source (or, sub-source) level propositions may not be an option if the needs relate to posited activities and factors such as transfer and persistence that require expert knowledge which needs to be taken into account in order to assess properly the findings in the context of the case as

a whole. Note that source level propositions may be adequate in cases when there is no risk that the court will misinterpret them in the context of the alleged activities in the case, that is, when no expert knowledge of factors such as transfer and persistence is needed. With transfer material, such as glass, this is rarely the case, as underlined by the ENFSI Guideline (ENFSI, 2015) (see Section 5.2).

The process of case pre-assessment can be summarised by the following steps:

- collection of task-relevant information that the scientist may need about the case;
- consideration of the issues in the case and the questions that the scientist can reasonably help address; this includes an assessment of the appropriate level of propositions under which the findings are to be evaluated;
- identification of the relevant factors that will appear in the likelihood ratio formula;
- assessment of the strength of the evidence, in terms of the likelihood ratio, that is expected given the background information;
- determination of the examination strategy;
- conducting analyses and gathering observations, leading to examination outcomes;
- calculation of the likelihood ratio for each of several potential findings.

An illustration of this procedure is presented in Section 5.5.3.

5.5.2 Pre-assessment of Evidence

Examples that consider the pre-assessment of various types of scientific evidence include the following. Cook et al. (1998b) discuss pre-assessment through the example of a hypothetical burglary involving potentially transferred glass fragments (i.e. an unknown quantity, Q, of recovered fragments). Stockton and Day (2001) consider an example involving signatures on questioned documents. Champod and Jackson (2001) consider a burglary case involving fibre evidence. Booth et al. (2002) discuss a drug case. A cross-transfer (or two-way transfer) case involving textiles is presented by Cook et al. (1999) where it is shown how pre-assessment can be updated when a staged approach is taken; the results of the examination of one of the garments are used to inform the decision about whether the second garment should be examined.

Puch-Solis and Smith (2002) describe a pre-assessment procedure within a training package to assess fibre evidence. The purpose is to provide support to forensic scientists to help determine whether an analysis of the collected fibres is cost effective. Possible values for the likelihood ratios, determined under the prosecution and defence propositions, are assigned to seven categories: (i) strong support for the defence, (ii) support for the defence, (iii) weak support for the defence, (iv) no support, (v) weak support for the prosecution, (vi) support for the prosecution, and (vii) strong

support for the prosecution. The scientist considers the probabilities for the likelihood ratios, given the proposition of the prosecution or defence. If there is a high probability for support for a particular proposition, then advice can be given to proceed with the analysis.

Theory and detailed examples are presented in Jackson et al. (2014). For a discussion on pre-assessment and interpretation, see Jackson et al. (2006), Jackson and Jones (2009), and Jackson (2013).

5.5.3 Pre-assessment: A Practical Example

In Section 5.5.1, the approach to pre-assessment proposed by Cook et al. (1998a,b) was presented from a general point of view. The aim of this section is to develop a practical example in a case involving textile fibres. Consider the following stages of pre-assessment as presented in Champod and Jackson (2001): identification of the task-relevant information the scientist may need, identification of the propositions used to assess the findings, progress through the pre-assessment of the case, determination of the examination strategy, assignment of the likelihood ratio and assessment of sensitivity, and identification of the effect of a change in the propositions.

5.5.3.1 Task-Relevant Information

Two armed and masked men burst into a post office. They threatened the staff, the takings for the day were handed over, and the men left. Witnesses said that one of the men was wearing a dark green balaclava mask and the other man was wearing a knotted stocking mask. They also said that the two men drove away from the scene in a car driven by a third man. Further along the presumed getaway route, a dark green balaclava was found. Mr U was arrested the following day. He denied all knowledge of the incident. Reference samples of his head hair and blood were taken as well as combings from his head hair. Mr U has not yet been charged with the robbery because there is very little evidence against him.

5.5.3.2 Formulation of Propositions of Interest and Events

Investigators are interested in knowing whether Mr U has worn the recovered mask because the intended charge is robbery. The scientist is, at first, able to define source-level propositions, such as

H_{p1}: Hairs in the mask came from Mr U;

H_{d1}: Hairs in the mask came from someone else.

H_{p2}: Saliva in the mask came from Mr U;

H_{d2}: Saliva in the mask came from someone else.

H_{p3}: Fibres in U's hair combings came from the mask;

H_{d3}: Fibres in U's hair combings came from some other garment or fabric.

The scientist may seek, however, to assess the findings given activity-level propositions (e.g. H_p, Mr U wore the mask at the time of the robbery, and H_d, Mr U has never worn the mask) because such propositions are, at a later point in the process, more relevant for the court and closer to the main issue in the case (Taroni et al., 2013). However, in order to work with such propositions, scientists will need to ensure that they (i) have task-relevant background information on the offence (e.g. time of the offence, time of arrest, time the traces are collected) to be able to assign the factors that relate to transfer and persistence of the recovered trace material of interest, and (ii) can collect data (i.e. literature) on transfer, persistence, and recovery of hairs, fibres, and saliva when someone wears a mask, as well as survey data on masks. Published data on hairs and saliva are very limited; data for fibres transfer to hair and for persistence are available though from more than 20 years ago. See, for example, Ashcroft et al. (1988), Salter and Cook (1996), and Cook et al. (1997).

If criteria (i) and (ii) are satisfied, the scientist can consider fibres first. If no sufficient background information is available to evaluate the findings (fibres, saliva, hairs) regarding

activity-level propositions, it will not be possible to offer an evaluative report that is helpful for the investigators or the court at the appropriate propositional level. Experts will have to limit their reporting to a statement of the analytical results. Assessment of the findings given source-level propositions can be prejudicial for the defendant if factors requiring expert knowledge, such as transfer and persistence, remain unaddressed and should not be left to the fact-finder (judge or jury) to consider. Thus, in the aforementioned case, the strategy would be to offer an assessment given activity-level propositions for fibres.

The second step in the pre-assessment is to determine the possible findings. Regarding fibres on hair combings, the scientist could consider the following results: no fibres are observed, a small number of fibres are observed (i.e. 1–3), or a large number of fibres are observed (i.e. more than 3). Note that the definitions of these categories is flexible and may depend on the available data or on the need to enhance the use of fibres to discriminate between the two propositions of interest. Note also that more than one group of fibres could be present, a version of the case that is not considered in this example.

Next, principal events involved in the activity level assessment are defined. To help the scientist determine which events are relevant for the pre-assessment of the findings in such a case, it is useful to consider questions of the kind 'What could happen if Mr U wore the mask at the time of

the robbery?'. If Mr U wore the mask, then three possibilities are as follows:

- *event* T_0: no fibres have been transferred, have persisted, and have been recovered;
- *event* T_s: a small number of fibres have been transferred, have persisted, and have been recovered;
- *event* T_l: a large number of fibres have been transferred, have persisted, and have been recovered.

Recovered fibres may be the consequence of either transfer during the course of the crime or presence beforehand by chance. The events linked to background presence are: no group of fibres is present by chance (event P_0); one group of fibres is present by chance (event P_1).

When a group of fibres is present by chance, it may be a small or a large group. The events are denoted as follows: the group of fibres present by chance is small (event S_s); the group of fibres present by chance is large (event S_l). Note that in Section 6.2.1, probabilities for events P_i and S_j are grouped using the notation $b_{g,\mathbf{m}}$.

Finally, consider that when a recovered group of fibres of unknown origin is compared with a control, two outcomes are possible: the recovered fibres correspond to (or cannot be distinguished from) the control with respect to the analysed features (event M); the recovered fibres do not correspond to (or can be distinguished from) the

control with respect to the analysed features (event $\bar{M}$). Note that when the scientist takes into account discrete features then, under the assumption that a sample known to come from a particular homogeneous source is compared with that source, the probability of observing corresponding features is often considered equal to 1.

5.5.3.3 Expected Likelihood Ratios

Upon analysing the hair combings, the scientist could observe one of the four situations given in Table 5.3. This list of outcomes does not take into account other possibilities such as the observation of a group of fibres coming from the transfer and a second group of fibres from the background. This aspect can be considered in a probabilistic graphical environment, such as Bayesian networks, as discussed in Taroni et al. (2014a).

To pursue the formulaic approach proposed here, clear assumptions are needed, such as for the number of groups. The analysis later considers that at most one group is present. The events occurring under the two activity-level propositions of interest, H_p and H_d as previously specified, are given in Table 5.4. This table also defines the probabilities related to the various events.

Note that the probabilities listed under H_p have a sum $1 - (t_s + t_l)p_1$, which is less than 1; the events to which they refer are not exhaustive. The probability $(t_s + t_l)p_1$ is the probability that two groups of fibres are transferred, one background group and one from the mask. The probabilities listed under

Table 5.3 Potential findings following the analysis of hair combings

Outcome	Number of groups	Number of non-corresponding groups	Number of corresponding groups	Size of matching groups
A	0	0	0	—
B	1	1	0	—
C	1	0	1	Small
D	1	0	1	Large

Table 5.4 Events and probabilities relating to findings under H_p and H_d

Outcome from Table 5.3	Events (Ev) to occur if H_p is true	Pr(Ev|H_p)	Events (Ev) to occur if H_d is true	Pr(Ev|H_d)
A	T_0,P_0	t_0p_0	P_0	p_0
B	T_0,P_1,$\bar{M}$	$t_0p_1(1-m)$	P_1,$\bar{M}$	$p_1(1-m)$
C	T_0,P_1,S_s,M orT_s,P_0	$t_0p_1s_sm + t_sp_0$	P_1,S_s,M	p_1s_sm
D	T_0,P_1,S_l,M orT_l,P_0	$t_0p_1s_lm + t_lp_0$	P_1,S_l,M	p_1s_lm

H_d do add to 1. The events to which they refer are exhaustive. Since it is assumed under H_d that there is no transfer of fibres from the mask, $t_s + t_l = 0$.

The next step in the pre-assessment process is to assign the various probabilities using data from published literature, case-specific experiments, or expert judgements based on the demonstrable and disclosable experience of the scientist. Consider the following:

- *Probabilities for events of transfer*: To assess the probabilities t_0, t_s, t_l, it is useful to answer questions like 'If the person of interest wore a mask, what is the probability that no/a small/a large number of fibres are transferred, persist and then would be recovered?' Note that information on the PoI (i.e. type and length of the PoI's hair), the material involved (i.e. sheddability), the methods used to search and collect fibres, and the circumstances of the case (i.e. alleged activities, time delays) are all relevant for a case-tailored assessment of probabilities.
- *Probabilities for background presence*: The probabilities p_0, p_1, s_s, s_l refer to the occurrence by chance of no fibres (p_0) or one group of fibres (p_1), which may be small (s_s) or large (s_l), as previously defined, on the hair in the event that no fibres have been transferred or if the PoI denied having worn the mask. Note that H_d specifies that the PoI never wore any mask. If the alternative proposition changes, for example, when the PoI states that they wore a

similar mask two days before the alleged facts, probabilities for background presence change, requiring new assessments.

- *Probabilities for observing corresponding features*: The probability m, often called the 'match probability', represents an assessment of the relative rarity of the extraneous fibres found on the head of a person incorrectly accused of wearing a mask. This assessment focuses on fibres that correspond, by chance, to the reference fibres coming from the mask. The relative rarity of fibres may be assessed in various ways. Scientists can refer to literature where fibres have been recovered on hairs of individuals and consider the relative proportions of fibres presenting the features of interest. They can also use so-called target fibre studies, stressing however that such studies offer different probabilities, notably probabilities for observing by chance one target group of fibres that match the control, written formally as $\Pr(P_1, S_s, M|H_d)$ and $\Pr(P_1, S_l, M|H_d)$. These probabilities are different from m. Databases can also be used, assuming that the potential source of fibres are hats, neckwear, bedding, and jumpers, so that the scientist will be able to assess the relative rate of occurrence of the corresponding fibres in this population.

Examples for probability assignments are discussed in Champod and Jackson (2001), leading to likelihood ratios summarised in Table 5.5.

Table 5.5 Likelihood ratios for the outcomes from Table 5.4 with $t_0 = 0.01$, $t_s = 0.04$, $t_l = 0.95$; $p_0 = 0.78$, $p_1 = 0.22$; $s_s = 0.92$, $s_l = 0.08$; $m = 0.05$, as proposed by Champod and Jackson (2001)

Outcome from Table 5.3	Number of non-corresponding groups	Number of corresponding groups	$\Pr(Ev \mid H_p)$	$\Pr(Ev \mid H_d)$	V
A	0	0	$t_0 p_0$	p_0	0.01
B	1	0	$t_0 p_1 (1 - m)$	$p_1 (1 - m)$	0.01
C	0	1 (small)	$t_0 p_1 s_s m + t_s p_0$	$p_1 s_s m$	3.09
D	0	1 (large)	$t_0 p_1 s_l m + t_l p_0$	$p_1 s_l m$	842.05

Likelihood ratios obtained in this pre-assessment for fibres help to answer to the question 'Is it useful to proceed with the analysis of the fibres?'. It has been shown that all situations offer a likelihood ratio different from the value of 1, that is, evidence with no evidential support; see Tables 5.3 and 5.4. If no fibres at all are recovered (outcome A), or if a group of fibres is recovered and this group does not correspond to the control object (outcome B), likelihood ratios supporting the alternative proposition are obtained. On the other hand, if one group, small or large, of fibres is recovered (outcomes C and D), and the group matches the control group, then a likelihood ratio greater than 1 is obtained, supporting the prosecution's proposition over the alternative proposition.

As previously noticed, pre-assessment can be applied as one aspect of practical decision-making in more sophisticated cases. Imagine a cross- or two-way transfer case. Pre-assessment can be updated when a staged approach is taken. For example, the victim's pullover is analysed first, then the PoI's pullover (Cook et al., 1999). The results of the examination of one of the garments are used to inform the decision about whether the second garment should be examined. As specified by Cook et al. (1999), it is easy to see how the principles could be extended to other kinds of cases. For example, if the crime involves the smashing of a sheet of glass and if clothing is submitted from a PoI, then the phased approach could be

applied to the order in which the garments are examined. If examination of the jacket reveals no glass, then how does formal pre-assessment inform the decision about examining the trousers or shoes? Examples are studied by Jackson et al. (2014). Another example, dealing with gun shot residue, is presented in Biedermann et al. (2009a).

Their model, using Bayesian networks, allows scientists to account for findings in terms of broad categories, such as 'no (0)', 'few (1–4)', 'some (5–8)', etc. particles, rather than a highly detailed description that would be required of a count variable.

These examples support the founding idea of pre-assessment, which '[…] was driven not only by serious questions raised about the quality of expert opinion but also by the growing requirement to manage limited forensic science resources in the most appropriate and cost-effective way' (Jackson and Jones, 2009, at p. 483).

5.5.3.4 The Relevant Population of Fibres

To clarify some of the remarks made in Section 5.5.3.3 on probabilities for the observation of corresponding features (so-called 'match probabilities'), consider the following example. An offender attempted to enter the rear of a house through a hole which they cut in a metal grille but failed when a security alarm went off. They left the scene. About ten minutes after the offence, a PoI wearing a red pullover was apprehended in the vicinity of the house following information

from an eyewitness who testified seeing a man wearing a red pullover running away from the scene. At the scene, a tuft of red fibre was found on the jagged end of one of the cut edges of the grille.

If the propositions from the prosecution and defence are that the fibres at the crime scene came from the PoI (H_p), and the fibres at the crime scene did not come from the PoI (H_d), respectively, following the argument of Section 5.3.1, the value of the evidence is given by

$$V = \frac{\Pr(y \mid x, H_p, I)}{\Pr(y \mid H_d, I)},$$

where y is the evidence of the red fibres on the grille and x is the evidence of fibres from the PoI's red pullover.

If H_d is true, the probability in the denominator is the probability of finding the characteristics of the tuft of fibre in the grille in a population of potential sources, called the relevant population. Assume that a survey has been made of the characteristics in a relevant population and the proportion of fibres with these characteristics is γ. As in Section 5.3.1.1, the value of the evidence is then

$$V = \frac{1}{\gamma},$$

where the relevant population is defined by H_d and I. Some considerations for the definition in this example are given by Champod and Taroni (2017).

- If the proposition H_d is that the PoI had never been present at the scene, then the relevant population is defined as that of red upper garments worn by burglars, accepting that the eyewitness had seen the burglar and was correct in the report that the burglar was wearing a red upper garment.
- If the proposition H_d is that the PoI has been correctly identified by the witness but had never been in contact with the grille, then the relevant population is defined by the potential sources of red fibres without any distinction in respect of the colour of the garment worn by the burglar.
- In the absence of an eyewitness, if the PoI has been apprehended because he was wearing a red pullover and the fibres found at the crime scene were red then the relevant population is defined as that of potential perpetrators wearing red garments. Further discussion of the implications of such a so-called search strategy is given in Section 6.1.8.1.
- In the absence of an eyewitness, if the PoI has been apprehended independently of the forensic attributes of the tuft, then the relevant population is defined by the potential perpetrators without any distinction in respect to the colour of the garment.

Analogously, an example of the influence of the defence's strategy for the definition of the relevant population is given in Robertson and Vignaux (1995b). Blood that did not belong to

a murdered victim was found at the scene of a crime in Auckland. An eyewitness saw a man of Maori appearance running away from the scene. Subsequently, a Maori person was arrested. The prosecution's proposition, H_p, is that this man is the criminal. The defence has two possible alternatives:

H_{d1}: the accused was the person seen running away but was not the murderer;

H_{d2}: the accused person was not the person seen running away.

Under H_{d1} there is no information about the murderer, so the murderer is to be considered as a randomly selected person in New Zealand. The scientist will then use the population as a whole as the relevant population for the value of γ. Under H_{d2} it is implicitly assumed the eyewitness saw the murderer and correctly identified their ethnicity; there is information about the murderer. He was of Maori appearance, so the murderer is to be considered as a randomly selected person of Maori appearance. The scientist will then use the Maori population as the relevant population for the value of γ.

5.5.3.5 The Defence and a Change in the Proposition

Consider a modification of the case of pre-assessment for fibres. Imagine that the evidence is not potential fibres recovered on the PoI's head,

but potential hairs found in the mask worn by the offender. Mr U was arrested, as said before, the following day. Instead of denying all knowledge of the recovered mask, he admitted the mask was his own mask and he lost it the day before the crime. Reference samples of his head hairs were taken, and a quantity n of head hairs are recovered from the mask by the forensic laboratory. Investigators are interested in knowing if Mr U has worn the mask *at the time* of the robbery. As a consequence, a question of interest is how the pre-assessment changes if the alternative proposition becomes 'Mr U has not worn the mask at the time of the robbery but the day before'.

Recall a general expression of the likelihood ratio (5.6) that presents the value for the evidence under an alternative (activity level) proposition specifying that the PoI did not wear the mask, but that an unknown person wore it.

$$V = \frac{b_0 t_n + b_{1,m}\gamma t_0}{b_0 \gamma t'_n + b_{1,m}\gamma t'_0}. \qquad (5.20)$$

If the PoI specifies that he had worn the mask the day before the crime, then the first term in the denominator of the likelihood ratio changes from $b_0\gamma t'_n$ to $b_0 t'_n$. The reason for this is that no uncertainty about the characteristics of the recovered hairs needs to be taken into account, as they are the hairs from the PoI. Note further that, following the discussion presented in Section 5.3.2.1, the likelihood ratio reduces to t_n/t'_n. It

will depend on the probability of transfer under the two propositions: the probability of transfer, persistence, and recovery of n hairs if the PoI wore the mask at the time of the crime, and the probability of transfer[3] of n hairs if the PoI wore the mask the day before the crime.

5.6 COMBINATION OF ITEMS OF EVIDENCE

In Section 2.9, Bayesian networks have been introduced with the aim of addressing uncertainties that affect inference based on results of various kinds of scientific examinations conducted on traces such as fibres, glass fragments, and biological material (DNA). In Section 5.3.2.5, the cross-transfer of material such as textile fibres during a criminal activity has been analysed and discussed. It represents a first example of the combination of results. The networks introduced in Section 5.4 provide valuable assistance in addressing some of a wide range of issues that affect a coherent evaluation of probative value.

[3]Note again that the term 'transfer' involves the properties of 'transfer', 'persistence' and 'recovery' that are always by convention considered as a single entity. All properties are necessary for their inclusion in the evaluation of evidence at activity-level. Examples of an extended way to evaluate the evidence by considering separately the previous properties is presented in Taroni et al. (2014a) using Bayesian networks. The network structure was inspired by Halliwell et al. (2003).

Existing probabilistic solutions proposed in the scientific literature may be used as a guide to elicit appropriate network structures (e.g. Garbolino and Taroni, 2002). By providing an explicit representation of the relevant variables together with their assumed dependence and independence properties, Bayesian networks have the potential to clarify the rationale behind a given probabilistic approach, in particular, with formulae for likelihood ratios. However, these formulae may attain critical levels of complexity, even for single items of evidence. One often needs to account for particular sources of uncertainty, related to phenomena such as transfer, persistence, and background presence. It may thus become increasingly difficult to structure probabilistic analyses properly and to discern the relevant variables as well as their relationships. If, in addition, several items of evidence need to be combined, then even further complications may be expected. In such situations, Bayesian networks may assist forensic scientists in constructing coherent, transparent, and defensible arguments (Juchli et al., 2012; Juchli, 2016) and so handle multiple sources of evidence even in legal cases (see, for sake of illustration Hepler et al. (2007); de Zoete and Sjerps (2018); Neil et al. (2019); Fenton et al. (2019); Graversen et al. (2019)).

Section 5.6.1, illustrates generic patterns of inference in combining evidence. So-called *dissonant* and *harmonious* evidences are described in Section 5.6.2 with the use of Bayesian networks.

More on the technical aspects can be found in Schum (2001). From a forensic science point of view examples are presented in Taroni et al. (2014a).

5.6.1 A Difficulty in Combining Evidence: The Problem of Conjunction

Unlike the evaluation of single items of scientific evidence, the joint evaluation of several distinct items of forensic evidence using formal methods has, to date, received occasional rather than systematic attention. To some extent, this is remarkable since forensic science typically requires consideration of multiple items of evidence. A complication that arises in such settings is that the application of probability theory to multiple items of evidence becomes increasingly complex, even for apparently simple questions. An example is the problem known as the 'difficulty of conjunction'.

The difficulty of conjunction is closely tied with the difference between the probability of the evidence and the probability of an explanatory proposition (Section 2.2). It thus also connects to the problem of posterior probabilities as an expression for the value of the evidence. Historically, the difficulty in combining evidence was the subject of a debate between Cohen (1977,1988) and Dawid (1987). In essence, the problem is described as follows: two items of evidence, when considered

individually, support a particular proposition, but when considered in combination, they seem to produce lower support. As an illustration, let E_1 and E_2 denote two distinct items of evidence. These shall be used to draw an inference concerning some proposition of interest, say, H for convenience. H has two possible outcomes H_p and H_d, denoting the propositions proposed by, respectively, the prosecution and the defence. For the purposes of this example, H_p and H_d are deemed to be mutually exclusive and exhaustive. Imagine further that some evaluator would give a probability of 0.7 for H_p given the occurrence of either E_1 or E_2, that is $\Pr(H_p \mid E_1) = \Pr(H_p \mid E_2) = 0.7$. The probability of interest is $\Pr(H_p \mid E_1, E_2)$.

If E_1 and E_2 are considered to be independent, given H_p or H_d, their joint probability can be written as the product of the individual conditional probabilities, that is, $\Pr(E_1, E_2 \mid H_p) = \Pr(E_1 \mid H_p) \times \Pr(E_2 \mid H_p)$. It is now tempting to believe that $\Pr(H_p \mid E_1, E_2)$ is obtained analogously, that is, by $\Pr(H_p \mid E_1, E_2) = \Pr(H_p \mid E_1) \times \Pr(H_p \mid E_2)$. The apparent contradictory result of this (incorrect) procedure is $0.7 \times 0.7 = 0.49$, which is less than the probability of H_p given either E_1 or E_2.

At this stage it is useful to consider Bayes' theorem. For two items of evidence, E_1 and E_2, and propositions H_p and H_d, the odds form of Bayes' theorem is

$$\frac{\Pr(H_p \mid E_1, E_2)}{\Pr(H_d \mid E_1, E_2)} = \frac{\Pr(E_1, E_2 \mid H_p)}{\Pr(E_1, E_2 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

or

$$\text{Posterior odds} = \text{likelihood ratio (V)} \times \text{prior odds}.$$

Assuming equal prior probabilities, $\Pr(H_p) = \Pr(H_d)$, the target probability $\Pr(H_p \mid E_1, E_2)$ is thus given by $V/(1+V)$. The likelihood ratio can be obtained as follows:

$$\begin{aligned} V &= \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(E_2 \mid H_p)}{\Pr(E_2 \mid H_d)} \\ &\quad \text{by independence of } E_1 \text{and } E_2, \\ &= \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)} \times \frac{\Pr(E_2 \mid H_p)}{\Pr(E_2 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)} \\ &\quad \text{since } \Pr(H_p) = \Pr(H_d) \text{so } \frac{\Pr(H_p)}{\Pr(H_d)} = 1; \\ &= \frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)} \times \frac{\Pr(H_p \mid E_2)}{\Pr(H_d \mid E_2)} \\ &= \frac{0.7}{0.3} \times \frac{0.7}{0.3} = \frac{0.49}{0.09}. \end{aligned}$$

From this, the probability of interest

$$\Pr(H_p \mid E_1, E_2) = \frac{V}{1+V} = \frac{0.49/0.09}{1 + 0.49/0.09} = 0.84,$$

which is greater than 0.7. Thus, under the stated assumptions, the combination of the two items of evidence yields a higher probability for H_p than when considered separately.

This example illustrates that in cases where two items of evidence are deemed to provide relevant information for the same pair of propositions, the

value of the two pieces of evidence in combination cannot readily be determined by the sole use of the posterior probabilities of the respective propositions, based on isolated considerations of the single items of evidence. This is also one of the reasons why scales of conclusions (see Section 2.4.6) based on posterior probabilities, as have been proposed, for example, in the field of shoe-mark analyses (Katterwe, 2003) or handwriting examination (Köller et al., 2004), are inadequate means for the assessment of scientific evidence (Taroni and Biedermann, 2005).

Such inferential impasses may be avoided by following established inferential procedures based on the likelihood ratio. Examples of reasoning problems unfortunately used to demonstrate the limitations of probability theory in legal reasoning can be found and discussed in de Zoete et al. (2019).

By focusing on a likelihood ratio (Section 2.3.1), one can successively add one item of evidence at a time and examine the probability of a proposition of interest, H, for example, given the available evidence, as described in (2.16) Section 2.4.1. The posterior odds after considering one item of evidence, E_1, for example, become the new prior odds for the following item of evidence, E_2, say. In a more formal notation one thus has, for propositions H_p and H_d:

$$\frac{\Pr(H_p)}{\Pr(H_d)} \times \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} = \frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)} \,. \qquad (5.21)$$

The term on the right-hand side of (5.21) represents the odds in favour of the proposition H_p given E_1. When E_2, a second item of evidence, becomes available, one may proceed as follows:

$$\frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)} \times \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} = \frac{\Pr(H_p \mid E_1, E_2)}{\Pr(H_d \mid E_1, E_2)} \,. \tag{5.22}$$

Here the posterior odds in favour of the proposition H_p incorporate knowledge about both items of evidence, E_1 and E_2. The likelihood ratio for E_2, shown in the center of (5.22), allows for a possible dependency of E_2 on E_1. More generally, this way of proceeding was concisely summarised by Lindley as 'Today's posterior is tomorrow's prior' (Lindley, 2000, at p. 301).

5.6.2 Generic Patterns of Inference in Combining Evidence

When reasoning about a scientific finding or result, two aspects are of interest: inferential direction and inferential force. Inferential direction informs about which proposition, compared with a given alternative, is favoured by the evidence. Inferential force expresses the strength of evidential support. Multiple items of evidence may exhibit various combinations of inferential direction and force. With respect to inferential directions, two situations can be distinguished. Either the inferential directions will point towards different propositions, or they point towards the

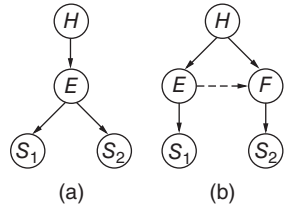

Figure 5.2 Generic Bayesian network structures for dissonant and harmonious evidence. Figure (a) accounts for situations of contradiction and corroboration and Figure (b) covers conflicting and converging evidence. The dotted arrow applies whenever one assumes a dependency between the two events E and F conditional upon H. In (a) S_1 denotes a source stating E did occur, S_2 denotes a source stating E did not occur. In (b), S_1 denotes a source saying E did occur, S_2 denotes a source saying F did occur.

same proposition. Following Schum (2001), the first situation involves evidence said to be 'dissonant', whereas the second situation involves evidence said to be 'harmonious'. Using various examples, the following sections discuss the probabilistic underpinnings of these distinctions in some further detail. Figure 5.2 represents the generic Bayesian network structures that underlie these patterns of reasoning.

5.6.2.1 Dissonant Evidence: Contradiction and Conflict

Contradiction All dissonant evidence incorporates an inferential divergence, although only some situations of dissonance can properly be called contradictory. Schum (2001) considered dissonant evidence that is not contradictory as being in 'conflict'. Properly speaking, a

contradiction is given only if the occurrence of mutually exclusive events is reported. In order to clarify this, consider a source S_1 stating that the event E occurred. Let this statement of S_1 be denoted by E^*. The asterisk in this notation is chosen to refer to a report about the occurrence of event E. This is different from the situation mentioned in Section 2.9.2.3, Example 2.9, concerning the assessment of the potential of error based on a distinction between a reported correspondence (i.e. 'match') and an actual correspondence. Next, suppose also a second source S_2 that states $\bar{E}^*$, that is, 'Event E did not occur'.

In a case involving questioned documents, it may be of interest to learn something about the event E, that the PoI wrote a signature on a handwritten document. Denote by $\bar{E}$ the alternative event that the PoI did not write the questioned signature, but that someone else did. One cannot directly know whether or not the PoI is the author of the questioned signature. One may therefore rely on an opinion presented by, for example, an eyewitness. Let this source of information be denoted by S_1. The report given by this source is written E^*, that is, a statement that E occurred. Next, one may also have a further source of information, given by source S_2. This source, too, reports on the proposition E, but affirms that its negation, $\bar{E}$, holds. An example for such a second source of information could be another eyewitness or a forensic document examiner, though it is emphasised that it is not appropriate for the latter to opine directly on source level propositions.

Given these outcomes, a question of interest then is how to draw an inference about a pair of ultimate propositions H_p and H_d, whilst allowing uncertainty about the event E now considered intermediate between E^* and H. Thus, $\Pr(E^* \mid E, H_p) = \Pr(E^* \mid E)$, $\Pr(\bar{E}^* \mid \bar{E}, H_p) = \Pr(\bar{E}^* \mid \bar{E})$, $\Pr(E^* \mid E, H_d) = \Pr(E^* \mid E)$ and $\Pr(\bar{E}^* \mid \bar{E}, H_d) = \Pr(\bar{E}^* \mid \bar{E})$. For the example introduced earlier, the proposition H could be, for example, the commission of a fraud, or any other criminal activity for which the establishment of authorship of the questioned signature at hand is inferentially relevant. For such a situation, the likelihood ratio for the two reports $\{E^*, \bar{E}^*\}$ takes the following form:

$$V_{E^*,\bar{E}^*} = \frac{\Pr(E^*, \bar{E}^* \mid H_p)}{\Pr(E^*, \bar{E}^* \mid H_d)} \,. \tag{5.23}$$

Assuming a dependency structure as shown in Figure 5.2a, the likelihood ratio in (5.23) can be presented in some further detail as follows (Schum, 2001):

$$\begin{aligned} V_{E^*,\bar{E}^*} &= \frac{\Pr(E^*, \bar{E}^* \mid H_p)}{\Pr(E^*, \bar{E}^* \mid H_d)} \\ &= \frac{\Pr(E \mid H_p) + \left[\frac{h_1 m_2}{f_1 c_2} - 1\right]^{-1}}{\Pr(E \mid H_d) + \left[\frac{h_1 m_2}{f_1 c_2} - 1\right]^{-1}} \,, \end{aligned} \tag{5.24}$$

where h_1 denotes $\Pr(E^* \mid E)$, m_2 denotes $\Pr(\bar{E}^* \mid E)$, f_1 denotes $\Pr(E^* \mid \bar{E})$, and c_2 denotes $\Pr(\bar{E}^* \mid \bar{E})$,

based on notation largely adopted from Schum (2001). The extended form of the likelihood ratio shown in (5.24) is reproduced here because it contains the expression $[(h_1 m_2 / f_1 c_2) - 1]^{-1}$. This part of the formula is also referred to as the *drag coefficient*, as it acts like a drag upon $V_E = \Pr(E|H_p)/\Pr(E|H_d)$ (Schum, 2001,2009). That is, if the likelihood ratio is thought as a force acting on the probabilities of H_p and H_d to move them from their prior positions to posterior positions, then the size of the force is measured by the value of V_E, and the drag coefficient acts as a drag on this force. The drag coefficient accounts for the credibility of the statements made by the sources of interest. In particular, it determines the closeness of $V_{E^*,\bar{E}^*}$ to V_E. Note that the likelihood ratio V_E describes the inference about H on the basis of the intermediate variable E and is given by the ratio of the two likelihoods $\Pr(E \mid H_p)$ and $\Pr(E \mid H_d)$.

The result shown in (5.24) can be further understood by considering local likelihood ratios for drawing an inference about E, on the basis of the distinct items of information $S_1 = E^*$ and $S_2 = \bar{E}^*$. In other words, one can write a likelihood ratio for item of evidence E^*, written V'_{E^*}, and one for the item of evidence $\bar{E}^*$, written $V'_{\bar{E}^*}$:

$$V'_{E^*} = \frac{\Pr(E^* \mid E)}{\Pr(E^* \mid \bar{E})} = \frac{h_1}{f_1}, \quad V'_{\bar{E}^*} = \frac{\Pr(\bar{E}^* \mid E)}{\Pr(\bar{E}^* \mid \bar{E})} = \frac{m_2}{c_2}.$$

A prime ($'$) is used here to indicate that the likelihood ratio concentrates on an inference

about E only, rather than about the ultimate proposition H.

When taking the reciprocal of the latter likelihood ratio, $V'_{\bar{E}^*}$, then one has an expression of the degree to which $\bar{E}^*$ favours $\bar{E}$: $V^{-1}_{\bar{E}^*} = \Pr(\bar{E}^* \mid \bar{E}) / \Pr(\bar{E}^* \mid E) = c_2/m_2$. It can now be seen that the overall support of the two statements $S_1 = E^*$ and $S_2 = \bar{E}^*$ for E depends on the relative magnitude of V'_{E^*} and $V'_{\bar{E}^*}$. In particular, in all the cases where $(h_1/f_1) > (c_2/m_2)$, the evaluation of the statements will strengthen the proposition E, in the sense that the support for E given E^* is greater than the support for $\bar{E}$ given $\bar{E}^*$. Conversely, if $(h_1/f_1) < (c_2/m_2)$, then the alternative proposition, $\bar{E}$, will be favoured.

More generally, notice further that the overall inferential force $V_{E^*,\bar{E}^*}$ is bounded by $V_{\bar{E}}$ and V_E so that $V_{\bar{E}} \leq V_{E^*,\bar{E}^*} \leq V_E$. That is, V_E represents the capacity of E to discriminate between H_p and H_d, given by $\Pr(E \mid H_p) / \Pr(E \mid H_d)$, whereas $\mathrm{V}_{\bar{E}}$ represents the discriminative capacity of $\bar{E}$, given by $\Pr(\bar{E} \mid H_p) / \Pr(\bar{E} \mid H_d)$. Usually, however, complete knowledge of the occurrence of either E or $\bar{E}$ will not be available, only evidence in the form of the statements $\{E^*, \bar{E}^*\}$.

Conflict Situations of evidence in conflict differ from those involving contradicting evidence because 'conflict' relates to events that are not mutually exclusive. This is illustrated in Figure 5.2b. For this model, suppose that source S_1 states E^*, that is, the occurrence of event E, which

is one that favours the proposition H_p. A second source, S_2, states F^*, that another event F, favouring proposition H_d, has occurred. The example given hereafter illustrates this outcome.

Consider again a report E^* that event E occurred, such as the event that a given PoI wrote a signature on a questioned document. Imagine further that the questioned document bears ridge skin marks (i.e. fingermarks). Let F denote the event according to which the fingermarks come from some person other than the PoI. Let F^* denote a scientist's report of such a conclusion. Again, it is emphasised that experts should not opine in this way, but it must realistically be conceded that such an approach represents currently the most widespread reporting practice, and this raises the question of how to process such a conclusion within a coherent framework. Conversely, let $\bar{F}$ denote the event according to which the fingermarks come from the PoI. Assuming that the fingermarks are found in a position on the questioned document where marks from the author of the crime of interest would be expected to be found, the event F can be considered relevant in an inference about the proposition H, that is, 'the person of interest is the author of the fraud'. Clearly, event F would favour H_d here because the probability of F can be reasonably be taken to be greater given H_d than given H_p. Stated otherwise, the likelihood ratio for F, written $V_F = \Pr(F \mid H_p) / \Pr(F \mid H_d)$, is smaller than 1. This represents support for H_d, compared with

H_p. In turn, the event E, which relates to the authorship of the questioned signature, provides support for H_p, rather than H_d. The likelihood ratio for E is $V_E = \Pr(E \mid H_p)/\Pr(E \mid H_d)$, and is greater than 1.

In this example, the evidential values of the reports E^* and F^* by, respectively, source S_1 and source S_2, are given by

$$V_{E^*} = \frac{\Pr(E^* \mid H_p)}{\Pr(E^* \mid H_d)} = \frac{\left(\begin{array}{c}\Pr(E^* \mid E)\Pr(E \mid H_p) \\ +\Pr(E^* \mid \bar{E})\Pr(\bar{E} \mid H_p)\end{array}\right)}{\left(\begin{array}{c}\Pr(E^* \mid E)\Pr(E \mid H_d) \\ +\Pr(E^* \mid \bar{E})\Pr(\bar{E} \mid H_d)\end{array}\right)}, \tag{5.25}$$

and

$$V_{F^*} = \frac{\Pr(F^* \mid H_p)}{\Pr(F^* \mid H_d)} = \frac{\left(\begin{array}{c}\Pr(F^* \mid F)\Pr(F \mid H_p) \\ +\Pr(F^* \mid \bar{F})\Pr(\bar{F} \mid H_p)\end{array}\right)}{\left(\begin{array}{c}\Pr(F^* \mid F)\Pr(F \mid H_d) \\ +\Pr(F^* \mid \bar{F})\Pr(\bar{F} \mid H_d)\end{array}\right)}, \tag{5.26}$$

where use is made of the results that $\Pr(E^* \mid E, H_p) = \Pr(E^* \mid E)$ and $\Pr(F^* \mid F, H_p) = \Pr(F^* \mid F)$ with similar results for E^* and $\bar{E}$ and F^* and $\bar{F}$. The two likelihood ratios (5.25) and (5.26) incorporate uncertainty about the actual – but unobserved – state of the events E and F, respectively. This is achieved by writing a given report, for example, E^*, conditioned on both E and $\bar{E}$,

weighted by the probability of, respectively, E and $\bar{E}$. Note further that the representation of the value of these two pieces of evidence, E^* and F^*, as the product of the separate likelihood ratios requires an assumption of conditional independence given H.

The two likelihood ratios (5.25) and (5.26) can also be written in a more compact form as follows (Schum, 2001):

$$V_{E^*} = \frac{\Pr(E \mid H_p) + \left[\frac{h_1}{f_1} - 1\right]^{-1}}{\Pr(E \mid H_d) + \left[\frac{h_1}{f_1} - 1\right]^{-1}}, \qquad (5.27)$$

$$V_{F^*} = \frac{\Pr(F \mid H_p) + \left[\frac{h_2}{f_2} - 1\right]^{-1}}{\Pr(F \mid H_d) + \left[\frac{h_2}{f_2} - 1\right]^{-1}} \qquad (5.28)$$

where $h_1 = \Pr(E^* \mid E)$, $f_1 = \Pr(E^* \mid \bar{E})$, $h_2 = \Pr(F^* \mid F)$, and $f_2 = \Pr(F^* \mid \bar{F})$. The fractions h_1/f_1 and h_2/f_2 represent the evidential values – that is the likelihood ratios – of the reports E^* and F^* for discriminating between the states of the distinct events E and F.

Given the assumption of conditional independence given H, the overall evidential value of the two reports E^* and F^*, that is, V_{E^*,F^*}, is given by the product of the individual likelihood ratios: $V_{E^*,F^*} = V_{E^*} \times V_{F^*}$. In the last example, such an assumption seems reasonable, notably it appears reasonable to consider ridge skin surface

characteristics as independent of handwriting characteristics.

More generally, if the events $\{E, \bar{E}\}$ and $\{F, \bar{F}\}$ need to be considered as not necessarily conditionally independent upon $\{H_p, H_d\}$, then the overall likelihood ratio will be of the form $V_{E^*} \times V_{F^*|E^*}$. The likelihood ratio for the second report F^* from source S_2 is then conditioned upon knowledge of the first report E^* from source S_1. More formally, this is written as $V_{F^*|E^*}$. Whilst V_{E^*} is as defined earlier in (5.25), the term $V_{F^*|E^*}$ involves a more extended development that can be shown to reduce to

$$V_{F^*|E^*} = \frac{\left(\begin{array}{c} \Pr(E \mid E^*, H_p)[\Pr(F \mid E, H_p) \\ -\Pr(F \mid \bar{E}, H_p)] \\ +\Pr(F \mid \bar{E}, H_p) + \left[\frac{h_2}{f_2} - 1\right]^{-1} \end{array} \right)}{\left(\begin{array}{c} \Pr(E \mid E^*, H_d)[\Pr(F \mid E, H_d) \\ -\Pr(F \mid \bar{E}, H_d)] \\ +\Pr(F \mid \bar{E}, H_d) + \left[\frac{h_2}{f_2} - 1\right]^{-1} \end{array} \right)}. \quad (5.29)$$

Here $h_2 = \Pr(F^* \mid F)$ and $f_2 = \Pr(F^* \mid \bar{F})$. These two terms represent, respectively, the numerator and denominator of a local likelihood ratio V'_{F^*} that expresses the degree to which the report F^* discriminates between the intermediate propositions F and $\bar{F}$.

There is a close relationship with respect to (5.28). In fact, when E is irrelevant for the assessment of F conditional on H, then (5.29) reduces

to (5.28). When knowledge of E is irrelevant, this relationship is expressed as

$$\Pr(F \mid E, H_p) = \Pr(F \mid \bar{E}, H_p) = \Pr(F \mid H_p) \text{ and}$$
$$\Pr(F \mid E, H_d) = \Pr(F \mid \bar{E}, H_d) = \Pr(F \mid H_d),$$

and this eliminates the terms $\Pr(E \mid E^*, H_p)[\Pr(F \mid E, H_p) - \Pr(F \mid \bar{E}, H_p)]$ and $\Pr(E \mid E^*, H_d)[\Pr(F \mid E, H_d) - \Pr(F \mid \bar{E}, H_d)]$ in, respectively, the numerator and the denominator of the likelihood ratio $\mathrm{V}_{F^* \mid E^*}$.

5.6.2.2 Harmonious Evidence: Corroboration and Convergence

Corroboration There are two main cases of harmonious evidence that can be distinguished, corroborating evidence and convergent evidence. The former case, corroboration, applies to evidence from sources that relate to the occurrence of the same event. As previously illustrated, consider two sources S_1 and S_2 both of which state E^*, that event E occurred. Suppose further that $\Pr(E \mid H_p) > \Pr(E \mid H_d)$, that is, event E is one that is more probable to occur if H_p is true, rather than when the specified alternative, H_d, is true. Using notation introduced so far, this expression of evidential value is written as V_E.

Consider, for example, a case where each of two independent handwriting experts reports E^*, that is, they provide a report of the event E, where E is defined as 'The person of interest is the source of the signature on the questioned document'.

These reports represent evidence from two distinct sources. In such a setting, each expert reports the occurrence of the same event E. In turn, E is relevant for an inference about H_p, the proposition according to which the PoI is the person who committed a given crime. In a Bayesian network, proposition H_p may be called an *ultimate probandum* because it is a root variable with no entering arcs from other nodes. In such a situation, H is said to be a graphical parent of node E.

When there is a dependence relationship between the variables as shown in Figure 5.2a, the likelihood ratio for the reports E_1^* and E_2^* by, respectively, sources S_1 and S_2, follows the general structure defined earlier in (5.23). For the case considered here, the expression can again be developed further and shown to be as follows:

$$\begin{aligned} V_{E_1^*,E_2^*} &= \frac{\Pr(E_1^*, E_2^* \mid H_p)}{\Pr(E_1^*, E_2^* \mid H_d)} \\ &= \frac{\Pr(E \mid H_p) + \left[\frac{h_1 h_2}{f_1 f_2} - 1\right]^{-1}}{\Pr(E \mid H_d) + \left[\frac{h_1 h_2}{f_1 f_2} - 1\right]^{-1}} . \end{aligned} \tag{5.30}$$

The overall inferential force of E_1^* and E_2^* does not only depend on the value of evidence E for the comparison of H_p and H_d, as expressed by the likelihoods $\Pr(E \mid H_p)$ and $\Pr(E \mid H_d)$. It also depends on the conditional probabilities of the reports given E, that is, the local likelihood

ratios $V'_{E_1^*} = h_1/f_1$ associated with the first report, and $V'_{E_2^*} = h_2/f_2$ associated with the second report.

Notice further that (5.30) can also be extended to multiple, say, n, independent sources, assuming conditional independence as before. For such a situation, the likelihood ratio can be shown to take the following form:

$$V_{E_1^*,\ldots,E_n^*} = \frac{\Pr(E_1^*, \ldots, E_n^* \mid H_p)}{\Pr(E_1^*, \ldots, E_n^* \mid H_d)} = \frac{\Pr(E \mid H_p) + \left[\prod_{i=1}^{n} \frac{h_i}{f_i} - 1\right]^{-1}}{\Pr(E \mid H_d) + \left[\prod_{i=1}^{n} \frac{h_i}{f_i} - 1\right]^{-1}}. \quad (5.31)$$

Such a setting is typically encountered in so-called testing cases, where n independent examiners work on a well-defined question; this could be an actual case or an experiment under predefined testing conditions (such as a proficiency test).

In order to obtain so-called overall corroboration with respect to the proposition H_p, it is necessary that $\prod_{I=1}^{n} \frac{h_i}{f_i} > 1$.

Notice further that the likelihood ratios in (5.30) and (5.31) cannot exceed V_E or $V_{\bar{E}}$. The joint value in an inference about H, based on a given number of individual sources that report on E, cannot be higher than that for complete knowledge about E, that is, a situation in which the actual state of E was known. Alternatively, it can be said that the

values of the individual reports for discrimination about H depend on the strength of the individual reports to discriminate between the states of the variable E. For example, if a report E^* is capable of making E certain, then the likelihood ratio for E^*, that is, V_{E^*}, would equate to that for E, that is V_E. However, if E^*, denoting multiple independent reports $E_1^*, \ldots, E_n^*$, leaves some uncertainty about E, then V_{E^*} would be less than V_E.

Convergence A situation of convergence is one in which two or more sources state the occurrence of distinct events that support different intermediate propositions that separately support the ultimate proposition. As depicted by Figure 5.2b, sources S_1 and S_2 may report the occurrence of the events E and F, which are conditionally independent given H. This is equivalent to having two independent lines of inference $E^* \rightarrow E \rightarrow H$ as illustrated in Figure 5.2(a) where S is substituted for E^* and the lines of inference are translated in the BN as $H \rightarrow E \rightarrow S$. In such a case, the overall likelihood ratio for the two reports E^* and F^* is given by the product of the likelihood ratios associated with the individual reports. That is, $V_{E^*,F^*} = V_{E^*} \times V_{F^*}$, and (5.25) and (5.26) can again be applied.

For illustration, suppose a scientist reports E^*, that event E occurred (e.g. a given PoI) wrote a signature on a questioned document. In addition, assume further that the questioned document bears ridge skin marks. Let F now denote the event that the fingermarks come from the PoI.

Hence, let F^* denote a scientist's report of such a conclusion. Assume that the fingermarks are found in a position on the document where marks from the offender would be expected to be found. The proposition F can then be considered relevant in an inference about the proposition H, that 'the person of interest is the person who committed the fraud'. The event F favours H when the probability of F is considered to be greater given H_p than given H_d. Stated otherwise, the likelihood ratio for F, written $V_F = \Pr(F \mid H_p)/\Pr(F \mid H_d)$, is greater than one. Given a likelihood ratio for E, written as $V_E = \Pr(E \mid H_p)/\Pr(E \mid H_d) > 1$, event F presents a further item in support of H, and thus implies convergence.

If, however, the events E and F are conditionally dependent upon the ultimate probandum H_p, then (5.29) is the relevant equation. In particular, in the assessment of the probative value of F, it is necessary to account for what has been observed in relation with the first item of evidence, E. This dependency is expressed by the conditional likelihood ratio $V_{F|E}$. According to the specified probabilistic underpinning, this may lead to the observation, known as *synergy*, that the second item, F, has more evidential value when E is already known, compared with a situation in which the outcome of the inspection of the first item of evidence is not known. In such a case, the evidence is called synergetic. However, it may also be the case that knowledge about E diminishes the inferential force of F. This would be a situation

sometimes referred to as redundancy. This may go as far as to entail a directional change, rather than only a reduction in the inferential force of F. An event F with $V_F > 1$ that supports H_p may, in the light of knowledge of E, lead to a situation in which $F \mid E$ supports H_d with $V_{F|E} < 1$.

Analyses of these and similar aspects of the joint evaluation of multiple items of evidence can be found in Lempert (1977), Biedermann and Taroni (2006), Juchli et al. (2012), Juchli (2016), Taroni et al. (2014a), de Zoete et al. (2015), Lucena-Molina et al. (2015a), de Zoete et al. (2017), and de Koeijer et al. (2019).

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

6

Evidence and Propositions: Practice

Chapter 5 developed technicalities for the derivation of equations for the value of the evidence measured in terms of the likelihood ratio. Equations for the assessment of transfer evidence given source, activity, and offence level propositions were presented with explanations on the way scientific literature reached these developments. This chapter presents examples of the use of these formulae.

6.1 EXAMPLES FOR EVALUATION GIVEN SOURCE LEVEL PROPOSITIONS

Start by considering, for the purpose of illustration, a simplified example with classical genetical

markers (i.e. the ABO system). Buckleton et al. (1987) provided data for blood group gene relative frequencies in New Zealand for different ethnic groups. These data may be used to derive numerical examples applicable to Sections 5.3.1.1 and 5.3.1.2. Consider the ABO blood grouping system and that the potential source (i.e. person of interest, PoI) and receptor (crime) stain are both of group B.

6.1.1 General Population

From Buckleton et al. (1987), the gene (relative) frequencies for New Zealand in this system are given in Table 6.1. Thus, for a crime for which the relevant population Ψ was the general New Zealand population,

$$V = 1/0.063 = 15.87 \simeq 16.$$

It is about 16 times more probable to observe this evidence if the PoI was the source of the recovered stain present at the crime scene than if they were not and an unknown person from the population Ψ was the source.

Table 6.1 Gene (relative) frequencies for New Zealand in the ABO system.

Blood group	A	B	O
Relative frequency	0.254	0.063	0.683

6.1.2 Particular Population

Suppose that background information I includes an eyewitness description of the criminal as Chinese. From Buckleton et al. (1987), the gene (relative) frequencies for Chinese in the ABO system are as given in Table 6.2. Then for a crime for which the Chinese population, Ψ_0, a subset of Ψ, was the relevant population,

$$V = 1/0.165 = 6.06 \simeq 6.$$

The evidence is six times more probable to be observed if the PoI was the source of the recovered stain at the crime scene than if they were not and an unknown person from the subpopulation Ψ_0 is the source. Thus the value of the evidence has been reduced by a factor of $15.87/6.06 = 2.62$ if there is external evidence that the criminal was Chinese.

Note that, as discussed in Section 5.3.1.4, in cases with material found on the crime scene (i.e. potentially left by the offender), evidence derived from blood group population data has to relate to the population from which the potential donors

Table 6.2 Gene (relative) frequencies for Chinese in New Zealand in the ABO system.

Blood group	A	B	O
Relative frequency	0.168	0.165	0.667

come, *not* to that of the PoI (though they may be the same). More details are presented in Section 6.1.3. In order to use blood group population data for Chinese, I has to contain information, such as eyewitness evidence, about the criminal. Kaye (1987a, 1993, 2004) commented on this point. Referring to *People v. Pizarro* (2003) and *People v. Prince* (2005), Kaye (2008b) also wrote:

> *The California Court of Appeal applied* Pizarro*, insisting that independent evidence had to establish the racial or ethnic identity of the perpetrator for any statistics on the frequency of a DNA genotype in that racial or ethnic group to be relevant. (p. 310)*

As discussed in Buckleton et al. (1987), the data on blood groups for the general population have been derived from a weighted average of the blood group frequencies in each ethnic group or subpopulation, which make up the population from which the criminal may be thought to have come. The weights are taken to be the proportion of each ethnic group in the general population. For a full discussion on this aspect, see also Kaye (2004, 2008a,b). Kaye (2004) wrote:

> *According to the Pizarro Court, when a crime has been committed by someone whose race or ethnicity is not known, presenting data on the relative frequency in various racial or ethnic groups of the type of DNA found at the scene of the crime is 'objectionable' because it is extraneous, potentially irrelevant and prejudicial. The Court reaches the same conclusion concerning statistics computed under alternative hypotheses about possible contributors to a mixed stain. (p. 211)*

The statistical argument is said to be 'objectionable' because – as reported in Kaye (2008b) – the prosecution failed to present substantial evidence to prove that the perpetrator was Caucasian, Hispanic, or Afro-American. A solution is to derive a weighted average as presented through an example in Kaye (2008a). Further, note that the previous example on the genetic marker is a simplified version of how scientists should compute the denominator of the likelihood ratio. In most of the cases, one cannot assume independence between the PoI's profile and the true donor's profile. This peculiarity should be considered in all cases involving biological material. Section 6.1.7 provides further discussion on this aspect.

6.1.3 A Note on The Appropriate Databases for Evaluation Given Source Level Propositions

It is important to analyse and to discuss issues that pertain to the choice of relevant databases or populations for assigning values to the components of evaluative likelihood ratio procedures when propositions of interest are at source level. A detailed analysis on how to choose relevant data for assigning a likelihood ratio with source level propositions is proposed in Champod et al. (2004). Their paper also outlined various forms of likelihood ratios that are suitable for evaluative assessments given source level propositions.

To start, consider some initial notation and definitions in a hypothetical shoemark case. In particular, let E_r and E_c denote the observations made on the mark recovered on the crime scene and on the test prints obtained from the PoI's shoes under controlled conditions, respectively. The assumption made here is that shoes with obviously different general patterns are left aside, and that they can safely be discarded from further consideration. Suppose also that the following pair of source level propositions are considered: H_p: The footwear mark found on the crime scene was made with the PoI's shoe; H_d: The footwear mark found on the scene was made with some other shoe.

Notice that, to facilitate the presentation, the discussion hereafter will refer to only one crime mark, taken as most representative, rather than several marks. It is assumed that, even though there may be several fragmentary marks with a comparable general pattern, these may be grouped if they are found in a 'logical sequence' (i.e. position), or form a 'trail', on which the offender is thought to have moved. It is conceded at this point that the wording 'was made' in the propositions defined earlier implies an action, so that one may be tempted to consider them activity level propositions. Let us notice, however, that for the kind of evidence considered here (i.e. marks on a floor), as well as its associated process of generation, the question of 'source' necessarily requires a well-defined action (i.e. the

process of walking). Putting this to one side, the stated propositions are interpreted as source level propositions here essentially because no factors (or phenomena) such as transfer, persistence, and recovery – typically required in genuine evaluations given activity level propositions (e.g. for glass and fibre evidence as in Chapter 5) – are considered in the setting discussed here.

Let I_c and I_p refer to the framework of circumstances relevant to the crime and the person of interest, respectively. Note that I_c and I_p can be seen as part of I_b as presented in Section 2.4.4. For example, I_c pertains to aspects of the crime under investigation (such as the number of offenders), whereas I_p accounts for relevant descriptors of the PoI (such as their occupation or their lifestyle); both sources of information may have an impact on the choice of databases. Considering that E_r and E_c refer to a given group of corresponding marks left by manufacturing features, the likelihood ratio can be written, in its general form, as follows:

$$V = \frac{\Pr(E_r, E_c \mid H_p, I_c, I_p)}{\Pr(E_r, E_c \mid H_d, I_c, I_p)}.$$

Given the proposition H_d, according to which the mark recovered on the crime scene was made by a shoe other than that of the PoI, the observations E_r and E_c can be regarded as independent so that the denominator can be rewritten as follows:

$$\Pr(E_r, E_c \mid H_d, I_c, I_p)$$

$$= \Pr(E_r \mid H_d, E_c, I_c, I_p) \times \Pr(E_c \mid H_d, I_c, I_p)$$
$$= \Pr(E_r \mid H_d, I_c, I_p) \times \Pr(E_c \mid H_d, I_c, I_p).$$

Moreover, given H_d, the observations on the crime mark E_r can be considered independent of information pertaining to the PoI, I_p. Conversely, observations on the test prints made by the PoI's shoes are not affected by information about the crime if the PoI's shoes are not the source of the crime mark. Therefore, the denominator can be further reduced to

$$\Pr(E_r \mid H_d, I_c) \times \Pr(E_c \mid H_d, I_p).$$

Next, consider the numerator of the likelihood ratio in some further detail. Regarding the mark's features originating from general sole pattern as considered in this setting, it seems reasonable to accept the idea that such features are reasonably stable in time, and that there were no events that could have caused substantial changes to such major shoe sole features. Such an assumption can be further supported, for example, if the time lapse between the commission of the crime and the seizure of the PoI's shoes is short. Thus, given knowledge of the characteristics of the PoI's shoes, these characteristics can be expected to be found in the crime mark and be recognised by the examiner as being in agreement with the features in the known comparison prints, if the PoI's shoe is in fact the source of the mark (proposition H_p).

Therefore, $\Pr(E_r \mid H_p, E_c, I_c, I_p) = 1$ so that the extended form of the numerator

$$\Pr(E_r, E_c \mid H_p, I_c, I_p) = \Pr(E_r \mid H_p, E_c, I_c, I_p) \times \Pr(E_c \mid H_p, I_c, I_p),$$

can be written more concisely as

$$\Pr(E_r, E_c \mid H_p, I_c, I_p) = \Pr(E_c \mid H_p, I_c, I_p).$$

Combining the results for the numerator and the denominator, one thus obtains the following likelihood ratio:

$$V = \frac{\Pr(E_r, E_c \mid H_p, I_c, I_p)}{\Pr(E_r, E_c \mid H_d, I_c, I_p)} = \frac{1}{\Pr(E_r \mid H_d, I_c)} \times \frac{\Pr(E_c \mid H_p, I_c, I_p)}{\Pr(E_c \mid H_d, I_p)}. \quad (6.1)$$

It is worth noting that the assumption that $\Pr(E_r \mid H_p, E_c, I_c, I_p) = 1$ is one that may not generally hold by default. When considerations are extended to particular acquired features, for instance, then it may not be taken as certain that a given configuration of features, seen in the print of known source, will necessarily be observed in a mark made by the shoe sole of interest. In fact, repeated applications over time of a given shoe sole may lead to different configurations of marks appearing in the traces produced. Due to factors such as pressure, direction of application, use or walking surface characteristics, some aspects may

preferentially be reproduced, rather than others. In a strict sense, such considerations could also be applied for more general aspects, such as sole pattern. Indeed, footwear marks are often fragmentary and the general pattern of a sole is rarely reproduced in its entirety. Therefore, such characteristics may also show up differently in repeated applications of a given shoe sole. In other words, the extent of reproduction of a given sole pattern is subject to variation. These considerations imply that the probability of observing a given configuration of marks is less than one. It should also be noted that the common statement according to which given footwear features are compatible or, in agreement, with features observed in a mark does *not* necessarily imply that $\Pr(E_r, E_c \mid H_p, I_c, I_p)$ should be 1. A judgement of 'compatible with' merely says that $\Pr(E_r, E_c \mid H_p, I_c, I_p)$ is not zero. In fact, as noted earlier, there is much reason to consider that this probability should take a value lower than one. A practical example of such reasoning is presented in Biedermann et al. (2012b).

The result of the aforementioned development is that there are three distinct probabilities that make up (6.1). As discussed in Champod et al. (2004), the assignment of values to these component terms requires distinct data collections: (i) a 'crime-related database' focusing on materials associated with crimes of the appropriate nature, to help with assigning $\Pr(E_r \mid H_d, I_c)$, (ii) an 'offender-related database' focusing on people who are known to have some kind of association with a crime scene of the relevant kind, to help

assigning $\Pr(E_c \mid H_p, I_c, I_p)$, and (iii) an 'innocent suspect database' covering data on individuals who are known not to have an association with a crime of the relevant kind but yet have come to the notice of the police, to help in the assignation of $\Pr(E_c \mid H_d, I_p)$.

The populations from which these databases are taken are all relevant to the evaluation of evidence. A general definition of a relevant population is the following.

Definition. A *relevant population* is the population defined by the combination of the proposition H_d proposed by the defence and the background information I.

This population is also called *suspect population* by some authors. The definition clarifies why it is not appropriate to use reference data on the population from which the suspect originated as presented in Section 5.3.1.3. The discussion on this argument was initiated by Kaye (1987a), Evett and Weir (1991), Lempert (1991), Wooley (1991), Weir and Evett (1992), and Weir (1992). Lempert (1993) clarified the point:

> *This analysis does not mean, however, that the use of a black database is, for example, appropriate if a white man is arrested for rape in a black ghetto. The suspect population consists of those who are plausible suspects given those factors that condition suspicion. If a rape victim claims her assailant was white, the police are not going to arrest a black-appearing man for the crime no matter how many black men would have been potential suspects absent information about the defendant's race.*

> *The suspect population will consist of white-appearing males, and the database used to estimate the uniqueness of a defendant's allele configuration should reflect that fact. (pp. 4–5)*

Moreover, Robertson and Vignaux (1995b) asked a simple question to deal correctly with the identification of the relevant population. They wrote:

> *[…] the accused's race is not relevant to the probability of obtaining the evidence under the alternative hypothesis if that is someone else left the mark. […] We then have the question 'what sort of someone else?' and the answer depends on what is known about the perpetrator, not the accused. (p. 149)*

An interesting aspect has been put forward by Lempert (1993) who focused on the role of relatives as members of the relevant population. He noticed that 'the suspect population, which is to say the group of people who plausibly might be suspected of having committed the crime, contains members of the same imbreed group'. (pp. 2–3) So, this suspect population have allele configurations across the loci tested with a relative high probability of matching that of the defendant. Robertson and Vignaux (1993a) extended further this idea and its implication for the value of evidence and for the probability of the proposition of interest. They wrote:

> *In some cases it will not be in the interest of the defendant to make such a claim [to be member of a restricted population such as relatives]. If, for example, the defendant claims*

> *to be a Pitcairn Islander and argues that it will be possible that the perpetrator is also a Pitcain Islander, the use of a specific Pitcain Island database might be justified. However, if there are only 50 Pitcairn Islanders in the country where the crime was committed, then the prior odds are dramatically reduced and the impact of the DNA evidence may even be increased. (p. 4)*

This aspect is emphasised again in Robertson and Vignaux (1995b) by stating:

> *Let us re-emphasise that although the value of the evidence is decreased if the alternative perpetrator is a brother, so is the pool of possible suspects. In fact, specifying a brother as the alternative may reduce the prior odds from 1 in several million to 1 in three or four. The combined effect of this and the blood evidence may even be to strengthen the case against the accused. (pp. 42–43)*

The choice of the relevant population has, therefore, an impact on the value of the evidence and on the case as a whole.

6.1.4 Two Trace Problem

The single trace case described in Section 6.1.3 can be extended to a case in which two bloodstains, or two groups of fibres (or, more generally two traces having different features), have been left at the crime scene (Evett, 1987b). Imagine the following case. A crime has been committed by two men, each of whom left a bloodstain at the crime scene. The stains are analysed, one is found to be of type Γ_1, the other of type Γ_2. Later,

as a result of information completely unrelated to the bloodstains, a single PoI is identified. For situations in which the PoI is selected through a database search, see Section 6.1.8.1. The PoI's blood is found to be of type Γ_1. It is assumed there is no evidence in the form of injuries. The scientific evidence is confined solely to the results of the blood analysis. The two propositions of interest are

H_p: the crime stains came from the PoI and one other man (so the PoI and the other man are the sources of the recovered stains);

H_d: the crime stains came from two other (unrelated) men.

The scientific evidence E consists of two components E_r and E_c, defined as follows:

- E_r: two crime stains with profiles Γ_1 and Γ_2;
- E_c: the PoI's blood is of type Γ_1 (a similar result follows if the PoI is of profile Γ_2).

The value, V, of the evidence is given by

$$\frac{\Pr(E_r \mid H_p, E_c, I)}{\Pr(E_r \mid H_d, I)}. \tag{6.2}$$

Note that if the bloodstain did not come from the PoI, their profile is generally considered not relevant for the assessment of the denominator (on this aspect, see comments made in Section 6.1.7).

The scientist knows that profiles Γ_1 and Γ_2 occur with probabilities γ_1 and γ_2, respectively, in some relevant population. Assume that I contains no information which may restrict the definition of

the relevant population to a subset of the general population.

Consider the numerator of (6.2) first. This is the probability that the two stains are of profiles Γ_1 and Γ_2, given

- that the PoI is the source of one of the crime stains;
- that the PoI is of profile Γ_1 and
- all other information, I: from the assumption above I implies that profile population proportions in the general population are relevant for the case at hand.

Assume that the two crime stains are independent pieces of evidence in the sense that knowledge of the profile of one of the stains does not influence the probability that the other will have a particular profile. Let the two criminals be denoted A and B. Further, E_r may be considered in terms of the following two mutually exclusive partitions:

- E_{r1}: A is of profile Γ_1, B is of profile Γ_2;
- E_{r2}: A is of profile Γ_2, B is of profile Γ_1.

Partitions E_{r1} and E_{r2} may be further subdivided using the assumption of independence. Thus $E_{r1} = (E_{r11}, E_{r12})$ where

- E_{r11}: A is of profile Γ_1;
- E_{r12}: B is of profile Γ_2.

Similarly, $E_{r2} = (E_{r21}, E_{r22})$ where

- E_{r21}: A is of profile Γ_2;
- E_{r22}: B is of profile Γ_1.

Thus, since E_{r1} and E_{r2} are mutually exclusive:

$$\begin{aligned}&\Pr(E_r \mid H_p, E_c, I)\\ &\quad = \Pr(E_{r1} \text{ or } E_{r2} \mid H_p, E_c, I)\\ &\quad = \Pr(E_{r1} \mid H_p, E_c, I) + \Pr(E_{r2} \mid H_p, E_c, I),\end{aligned}$$

following (1.5), the second law of probability for mutually exclusive events. However, only one of these two probabilities is non-zero. If the PoI is A then the latter probability is zero; if the PoI is B then the former probability is zero. Assume, again without loss of generality, that the PoI is A. Then

$$\begin{aligned}&\Pr(E_r \mid H_p, E_c, I)\\ &\quad = \Pr(E_{r1} \mid H_p, E_c, I)\\ &\quad = \Pr(E_{r11} \mid H_p, E_c, I) \times \Pr(E_{r12} \mid H_p, E_c, I)\end{aligned}$$

by independence.

Now, $\Pr(E_{r11} \mid H_p, E_c, I) = 1$ since if the PoI was the source of one of the crime stains, and his profile is Γ_1, then it is certain that one of the profiles is Γ_1.

Also, $\Pr(E_{r12} \mid H_p, E_c, I) = \gamma_2$ since the second bloodstain was left by the other donor. At present, in the absence of information from I, the offender B is considered a member of the relevant population. The probability is thus assumed to be the population proportion of profile Γ_2 in the relevant population and this is γ_2. The numerator then takes the value γ_2.

Now consider the denominator of (6.2),

$$\begin{aligned}\Pr(E_r \mid H_d, I) &= \Pr(E_{r1} \mid H_d, I) + \Pr(E_{r2} \mid H_d, I)\\ &= \Pr(E_{r11} \mid H_d, I) \Pr(E_{r12} \mid H_d, I)\end{aligned}$$

$$
\begin{aligned}
&\quad + \Pr(E_{r21} \mid H_d, I) \Pr(E_{r22} \mid H_d, I) \\
&= \gamma_1\gamma_2 + \gamma_2\gamma_1 \\
&= 2\gamma_1\gamma_2.
\end{aligned}
$$

Thus,

$$V = \gamma_2/(2\gamma_1\gamma_2) = 1/(2\gamma_1). \qquad (6.3)$$

This result should be compared with the result for the single trace case, $V = 1/\gamma$. The likelihood ratio in the two-trace case is one half of what it is in the corresponding single trace case. This is intuitively reasonable. If there are two donors and one PoI, one would not expect the evidence of a matching bloodstain to be as valuable as in the case in which there is one donor and one PoI. Note that if γ_1 is greater than 0.5 then V is less than 1. The evidence is supportive of the proposition that the crime stain came from two men, not including the PoI. As has been commented, 'this appears counterintuitive at first sight but, rather than demonstrating a flaw in the logic, it demonstrates that intuition can be unreliable!' (Evett, 1990). An illustrative example, using an idea from the tossing of two coins, is given by Evett and Weir (1998, p. 34).

Other pairs of propositions may appear more appropriate in different situations. For example, suppose the stains are located at distinct positions on the crime scene. It could be that one was found on a carpet and another on a pillowcase. The propositions may then be

H_p : the PoI left the stain on the carpet;

H_d : two unknown people left the two stains *or*

H_p : the PoI left the stain on the carpet;

H_d : an unknown person left the stain on the carpet and another unknown person left the stain on the pillowcase.

The context of the case or the strategies of the prosecution and defence can influence the choice of the propositions. Different propositions could lead to different values for the evidence. It is necessary then to look at Bayes' theorem (2.12) in its entirety. Different propositions may have different prior odds $\Pr(H_p \mid I)/\Pr(H_d \mid I)$ as well as different values of the evidence. It follows that the posterior odds $\Pr(H_p \mid E, I)/\Pr(H_d \mid E, I)$ may also be different for different propositions. Further comments on these ideas may be found in Meester and Sjerps (2003). A detailed analysis with the aim of clarifying the interrelationships that exist amongst the different solutions (and in this way, produce a global vision of the problem) is presented in Gittelson et al. (2013a).

Further discussion of the value of evidence in the two trace problem when considering activity or crime level propositions illustrates that several factors have to be considered: the number of reported perpetrators, the relevance of each stain and the specification of transfer probabilities (Triggs and Buckleton, 2003). Gittelson et al. (2012a) developed a model to relax assumptions made by Triggs and Buckleton (2003) for the development of the likelihood ratio, generalised to cases involving n traces. Because the algebraic

approach becomes increasingly challenging to follow in cases with an increasing number of variables and dependence relationships between these variables, probabilistic graphical models (i.e. Bayesian networks) are suggested as a way to support the application of a likelihood ratio based evaluation. An example is presented in De March et al. (2016).

An extension to cases with n stains, k groups of stains, and m donors is given in Section 6.1.5, but without consideration of relevance or probabilities of transfer.

6.1.5 Many Samples

6.1.5.1 Many Different Profiles

Consider a crime in which n bloodstains are left at the crime scene, one from each of n donors. A single PoI is identified whose DNA profile corresponds to that of one of the bloodstains at the crime scene. Assume throughout that I contains no relevant information. Whilst hypothetical the example illustrates points that require consideration in the evaluation of evidence. A similar argument can be developed for a case in which there are n groups of fibres from n distinct sources and a PoI has been identified in whose possession is clothing whose fibres correspond to one of the groups.

Assume that the n bloodstains, one from each of n donors, all have different profiles. The two propositions to be considered are:

H_p: the trace from the crime scene came from the PoI and $(n-1)$ other people;

H_d: the trace from the crime scene came from n unknown people.

The scientific evidence E is defined as follows:

- E_r: the crime stains have profiles $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$,
- E_c: the PoI's profile is Γ_1.

The population proportions for the profiles $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ are, respectively, $\gamma_1, \gamma_2, \ldots, \gamma_n$.

Consider the numerator $\Pr(E_r \mid H_p, E_c, I)$ first. The PoI's profile corresponds to that of the stain of profile Γ_1. There are $(n-1)$ other donors who can be allocated to the $(n-1)$ other stains in $(n-1)!$ ways. Thus:

$$\begin{aligned}\Pr(E_r \mid H_p, E_c, I) &= (n-1)! \prod_{i=2}^{n} \gamma_i \\ &= (n-1)!(\gamma_2 \cdots \gamma_n).\end{aligned}$$

Now, consider the denominator. There are $n!$ ways in which the n donors, of whom the PoI is not one, can be allocated to the n stains. Thus:

$$\Pr(E_r \mid H_d, I) = n! \prod_{i=1}^{n} \gamma_i = n!(\gamma_1 \cdots \gamma_n).$$

Hence

$$V = \frac{(n-1)! \prod_{i=2}^{n} \gamma_i}{n! \prod_{i=1}^{n} \gamma_i} = \frac{1}{n\gamma_1}. \qquad (6.4)$$

6.1.5.2 General Cases

n Stains, k Groups, k Donors Suppose now that there are k different profiles $\Gamma_1, \ldots, \Gamma_k$, with population proportions $\gamma_1, \ldots, \gamma_k$ and that these profiles correspond to the profiles of k different people amongst the n stains ($k < n$) recovered on the crime scene and that the PoI has one of these profiles. Two propositions of interest to considered could be:

H_p: the stains from the crime scene came from the PoI and $(k-1)$ other people;

H_d: the stains from the crime scene came from k unknown people.

The scientific evidence consists of:

- E_r: the crime stains with profiles $\Gamma_1, \ldots, \Gamma_k$ and there are $s_1, \ldots, s_k$ ($\sum_{i=1}^{k} s_i = n$) of each,
- E_c: the PoI's profile is Γ_1 (without loss of generality).

The probabilities given later are in the form of the multinomial distribution (Section A.2.4). Consider the numerator $\Pr(E_r \mid H_p, E_c, I)$ first. The PoI's profile corresponds to that of the stains of profile Γ_1. There are $(n - s_1)$ other bloodstains that can be allocated in $(n-s_1)!/(s_2! \cdots s_k!)$ ways, where $\sum_{i=2}^{k} s_i = n - s_1$, to give

$$\Pr(E_r \mid H_p, E_c, I) = \frac{(n-s_1)!}{s_2! \cdots s_k!} \gamma_2^{s_2} \cdots \gamma_k^{s_k}.$$

Next, consider the denominator. There are $n!/(s_1!s_2!\cdots s_k!)$ ways in which the n stains, none of which is associated with the PoI, can be allocated to the profiles. Thus

$$\Pr(E_r \mid H_d, I) = \frac{n!}{s_1!\cdots s_k!}\gamma_1^{s_1}\cdots\gamma_k^{s_k}.$$

Hence

$$V = \frac{(n-s_1)!s_1!}{n!\gamma_1^{s_1}} = \frac{1}{\binom{n}{s_1}\gamma_1^{s_1}}. \qquad (6.5)$$

Notice that V is independent of k, the number of donors and that if $s_1 = 1$ the result reduces to that of (6.4).

n Stains, k Groups, m Donors A similar result may be obtained in the following situation. There are n bloodstains with k different profiles with s_i in the ith group ($\sum_{i=1}^{k} s_i = n$). There are m donors with m_i in each profile ($\sum_{i=1}^{k} m_i = m$) such that s_{ij} ($j = 1, \ldots, m_i$) denotes the number of stains belonging to the jth donor in the ith group and $\sum_{j=1}^{m_i} s_{ij} = s_i$ when it is assumed, without loss of generality, that the first set of stains in the first group came from the PoI. The denominator equals

$$\frac{n!}{s_{11}!\cdots s_{km_k}!}\gamma_1^{s_1}\cdots\gamma_k^{s_k}.$$

The numerator equals

$$\frac{(n-s_{11})!}{s_{12}!\cdots s_{km_k}!}\gamma_1^{s_1-s_{11}}\gamma_2^{s_k}\cdots\gamma_k^{s_k}.$$

Then

$$V = \frac{(n - s_{11})!s_{11}!}{n!} \times \frac{1}{\gamma_1^{s_{11}}} = \frac{1}{\binom{n}{s_{11}} \gamma_1^{s_{11}}},$$

a result similar to (6.5).

6.1.6 Multiple Propositions

Remember the principle of evidence evaluation that stipulates that for the assessment of the value of any item of scientific evidence, the forensic scientist needs at least two propositions. However, this raises the question of what to do if there are more than two propositions that should be considered. Typically, this happens in DNA evidence cases when relatives of the PoI are amongst the potential sources of the crime stain (Lempert, 1991; Buckleton and Triggs, 2005). Historically, two approaches have been proposed. The first approach focuses on the calculation of the posterior probabilities of the propositions. The second approach proposes the development of a Bayes' factor (see Section 2.3.2) for the multiple propositions. Below, the two approaches are presented using examples.

6.1.6.1 Multiple Propositions: Posterior Probabilities

Consider a situation in which the evidence E is that of a DNA profile of a stain of body fluid found at the scene of a crime and of the DNA profile from a PoI

which is reported to correspond to the profile of the crime stain. Further, suppose that there are three propositions to be considered:

H_p: the PoI is the donor of the crime stain;

H_{d1}: an unknown person from the population is the donor of the crime stain;

H_{d2}: a brother of the PoI is the donor of the crime stain.

This situation has been discussed by Evett (1992). Evett's original notation is retained, so that θ_0, θ_1, and θ_2 denote the prior probabilities for each of these three propositions ($\theta_0 + \theta_1 + \theta_2 = 1$). Assume that $\Pr(E \mid H_p) = 1$. Denote $\Pr(E \mid H_{d1})$ by ϕ_1 and $\Pr(E \mid H_{d2})$ by ϕ_2. Also, H_d, the complement of H_p, is assumed to be the conjunction of H_{d1} and H_{d2}. Then, using Bayes' theorem (2.4) and the law of total probability (1.12),

$$\begin{aligned}\Pr(H_p \mid E) &= \frac{\Pr(E \mid H_p)\theta_0}{\Pr(E \mid H_p)\theta_0 + \Pr(E \mid H_{d1})\theta_1 + \Pr(E \mid H_{d2})\theta_2} \\ &= \frac{\theta_0}{\theta_0 + \phi_1\theta_1 + \phi_2\theta_2}, \qquad (6.6) \\ \Pr(H_d \mid E) &= \frac{\phi_1\theta_1 + \phi_2\theta_2}{\theta_0 + \phi_1\theta_1 + \phi_2\theta_2},\end{aligned}$$

and hence the posterior odds in favour of H_p are

$$\frac{\theta_0}{\phi_1\theta_1 + \phi_2\theta_2}.$$

Let the relevant population size be N and let the number of siblings be n where $N \gg n$. It can be assumed that $\theta_0 = 1/N, \theta_1 = (N-n)/N$ and $\theta_2 = (n-1)/N$. The posterior odds in favour of H_p are approximately equal to

$$\frac{1}{\phi_1 N + \phi_2 (n-1)},$$

replacing $(N-n)$ with N.

Consider an example where the probability (ϕ_1) of a correspondence with a random person in the relevant population is assessed as $1/25\,000\,000$. If $N = 100\,000$, and there are no siblings so that $n = 1$ then the posterior odds in favour of H_p are $1/(N\phi_1) = 250$. If there is a brother, then $\phi_2 \simeq (1/4)^2$, $n = 2$ and the posterior odds are

$$1/\left\{\frac{1}{250} + \left(\frac{1}{4}\right)^2\right\} = \frac{1}{0.004 + 0.063} \simeq 15.$$

The existence of the brother has reduced the posterior odds by a factor of over 15 from 250 to 15. Note that ϕ_2 can be known given the PoI profile (e.g. $\Pr(G_2 \mid G_1)$, where G_1 and G_2 denote the genotypes of both individuals (Balding and Nichols, 1994). Bayesian networks are a simple tool to infer such a conditional probability.

A more general approach involving a brother is discussed by Balding (1997, 2000). The existence of just one brother amongst the possible donors of the crime stain can outweigh the effect of very many unrelated men for realistic values of the

profile probabilities. The possible crime stain donors are the PoI, one brother of the PoI (note that the brother may, however, be missing, or refuse to collaborate with investigators, or it may not even be known whether or not the PoI has any brothers) and one hundred unrelated men. Only the DNA profile of the PoI is available. Consider that the profile probability for the brother is 1/100 and the profile probability for the other men is 1/1 000 000. Suppose that the probability $\Pr(I \mid H_p)$ of the non-DNA evidence I was the same for all the possible donors of the crime stain (102 individuals). Then

$$\begin{aligned}\Pr(H_p \mid E, I) &= 1/\{1 + 0.01 + (100 \times 0.000001)\} \\ &= 0.99.\end{aligned}$$

Thus $\Pr(H_d \mid E, I) = 0.01$.

If the brother is ignored, then

$$\begin{aligned}\Pr(H_p \mid E, I) &= 1/\{1 + (100 \times 0.000001)\} \\ &= 0.9999\end{aligned}$$

and $\Pr(H_d \mid E, I) = 0.0001$. Consideration of the presence of the brother has increased the probability of the defence proposition by a factor of 100. See Section 6.1.8.3 for further details.

The consequences are more dramatic if other brothers, cousins, or other relatives are members of the relevant population of potential sources of the crime stain. It may in some cases be plausible that the non-DNA evidence has approximately

the same value for the PoI as for some or all of the siblings. In this case, it can be assumed that the probability, $\Pr(H_p \mid E, I)$, that the PoI is the donor of the stain is at most $1/(1 + qx)$ where x denotes the number of such siblings and q represents the sibling profile probability (Balding and Donnelly, 1995b).

Buckleton and Triggs (2005) raised a related question, whether a scientist should systematically consider (at least) two alternatives propositions in the denominator of Bayes' factor. The authors emphasised that it would seem reasonable to include this number if it is important for the decision making process. They remarked that:

> *The central thesis of this paper, then, is whether this match probability should be that for an unrelated person or for both the unrelated person and a brother. To do this we examine what contributes to the 'remainder' of the posterior probability that is not assigned to the PoI. This remainder represents the doubt, if any. It must fall on either the sibling or on one or more unrelated males. (p. 117)*

In the authors' analysis, the value for the posterior probability related to the 'brother' proposition increases as the number of genetic markers increases. Thus, there is the need to consider at least two propositions for the defence side.

An alternative development has been proposed in Evett and Weir (1998) referring to Balding and Donnelly (1995a,b) . Consider a scenario where the DNA profile of a recovered stain corresponds to that of a person of interest (PoI). Consider

also, as previously stated, that $\Pr(H_p \mid I) = \pi_1$ is the prior probability that the PoI is the source of the recovered stain and that $\pi_i (i = 2, \ldots, n)$ are the prior probabilities of the alternative propositions for other members of the relevant population: $\sum_i \Pr(H_{di} \mid I) = \sum_{i=2}^{n} \pi_i = 1 - \pi_1$. Assume that the probability of observing the recovered profile E_r given that the PoI has profile E_c and that they are the source of the recovered stain (H_p), $\Pr(E_r \mid E_c, H_p, I) = 1$, the posterior probability of proposition H_p is

$$\begin{aligned} \Pr(H_p \mid E_r, E_c, I) &= \frac{\pi_1}{\sum_{i=1}^{n} \Pr(E_r \mid E_c, H_{di}, I)\pi_i} \\ &= \frac{1}{1 + \sum_{i=2}^{n} \frac{\Pr(E_r \mid E_c, H_{di}, I)\pi_i}{\pi_1}}. \end{aligned} \tag{6.7}$$

The use of the ratio $w_i = \pi_i / \pi_1$ of the prior probabilities is suggested (Evett and Weir, 1998) as a 'weighting function that expresses how much more (or less) probable than the PoI is the *i*-th person to have left the crime stain, based on the non-DNA evidence'. (p. 41). Balding (2000) commented on w_i and further details are presented in Section 6.1.8.3. Therefore, the posterior probability on H_p becomes:

$$\Pr(H_p \mid E_r, E_c, I) = \frac{1}{1 + \sum_{i=2}^{n} \Pr(E_r \mid E_c, H_{di}, I) w_i}.$$

Another way to cope with multiple propositions are Bayesian networks. Details are presented

in Biedermann et al. (2012d) and Taroni et al. (2014a).

Note also that the consideration of three propositions has led to consideration of posterior odds in favour of one of the propositions. For comparing more than two propositions, propositions have to be combined in some meaningful way to provide two propositions for comparison. Both methods (posterior probabilities and Bayes' factors) have been applied in questioned document cases; for examples and discussions, see Köller et al. (2004) and Taroni and Biedermann (2005).

6.1.6.2 Multiple Propositions: Bayes' Factor

Consider the case of *State v. Klindt* (1968) discussed in Lenth (1986). This case has two aspects of interest. It illustrates a method for combining evidence and it illustrates a method for evaluating the evidence for more than two propositions. The example uses blood grouping whereas DNA profiling would now be used. However, the principles remain the same and this example provides a good example of them.

The case involved the identity of a portion of a woman's body. The portion was analysed and it was determined that the woman was white, aged between 27 and 40, had given birth to at least one child and had not been surgically sterilised. Also, the results for seven genetic markers were obtained. All but four missing persons in the area of the four states round where the body was

found were eliminated as possible identities for the woman. Label these four persons as P, Q, R, and S. Women Q, R, and S had been missing for six months, six years, and seven years, respectively, and their last known locations were at least 200 miles from where the body was found. Woman P had been missing for one month at the time the body was discovered and had last been seen in the same area.

The blood type of P was known to be A. The blood types of Q, R, and S were not known. The genetic constitution for the other markers was unknown for all four women. For the remaining six phenotypes, samples of tissues from the parents of P enabled a value of 0.5 to be calculated for the probability that the woman whose body was found had the phenotypes it had if it were that of P; i.e. Pr(given phenotypes| P) equals 0.5. See also Section 6.3.3 for another such example of kinship analysis. For some general population the proportion of individuals with these phenotypes was assigned as 0.00764. No kinship analyses were conducted for Q, R, or S. The ages of all four women were known and they were all mothers. However, in order to illustrate the procedure for combining evidence, Lenth (1986) made the following alterations to the actual data: it was not known whether Q was a mother; R's age was taken to be unknown and S was known to have type A blood.

The previous information is summarised in Table 6.3 using information on population

Table 6.3 Probability evidence in *State v. Klindt* (altered for illustrative purposes).

Indicator	Attribute	Woman, X			
		P	Q	R	S
	Age (years)	33	27	unknown	37
E_1	Mother|age	1.0	0.583	1.0	1.0
E_2	Sterilised| mother, age	1.0	0.839	0.662	0.542
E_3	Type *A* blood	1.0	0.362	0.362	1.0
E_4	Other six phenotypes	0.500	0.007 64	0.007 64	0.007 64
$\Pr(E_1, E_2, E_3, E_4 \mid \text{age}, X)$		0.5000	0.001 35	0.001 83	0.004 14
Fraction of total, $\Pr(X \mid E, \text{age})$		0.9856	0.0027	0.0036	0.0082

Source: From Lenth (1986). Reprinted with permissions of Elsevier.

proportions given by Lenth (1986). If a particular characteristic is known to hold for a particular woman, then a probability of 1 (corresponding to certainty) is entered. If the presence or absence of an attribute is not known, the population proportion of the attribute amongst the general population is given. There are four propositions to be compared, one for each of the four women whose body may have been found. The evidence to be assessed is $E = \{E_1, \ldots, E_4\}$, the four separate items of evidence listed in Table 6.3. The probabilities to be determined are of the form $\Pr(E_1 \cdots E_4 \mid \text{age}, X)$ where X is one of $P, Q, R,$ or S. These items of evidence are not all mutually independent. Denote the evidence as follows:

- E_1: mother, yes or no;
- E_2: not sterilised, yes or no;
- E_3: type A blood, yes or no;
- E_4: other six phenotypes.

Using the third law of probability for dependent events (1.9)

$$\begin{aligned}\Pr(E \mid \text{age}, X) &= \Pr(E_1, E_2, E_3, E_4 \mid \text{age}, X)\\ &= \Pr(E_1 \mid \text{age}, X) \times \Pr(E_2 \mid E_1, \text{age}, X)\\ &\quad \times \Pr(E_3 \mid E_2, E_1, \text{age}, X)\\ &\quad \times \Pr(E_4 \mid E_3, E_2, E_1, \text{age}, X)\end{aligned}$$

$$= \Pr(E_1 \mid \text{age}, X) \times \Pr(E_2 \mid E_1, \text{age}, X) \times \Pr(E_3 \mid X) \Pr(E_4 \mid X)$$

where X is one of P, Q, R, or S, and the final expression depends on certain independence relationships amongst the four pieces of evidence. For example, the probability of sterilisation is dependent on age and being a mother or not, whilst the probability an individual has type A blood (E_3) is independent of age, whether a mother or not (E_1), and whether or not sterilised (E_2). The probabilities for the combined evidence E are given in the penultimate row of Table 6.3, namely,

- $\Pr(E \mid \text{age}, P) = 0.5000$;
- $\Pr(E \mid \text{age}, Q) = 0.00135$;
- $\Pr(E \mid \text{age}, R) = 0.00183$;
- $\Pr(E \mid \text{age}, S) = 0.00414$.

Explicit dependence on 'age' is now omitted for ease of notation. The posterior probability that the body is that of P, $\Pr(P \mid E)$, may be assigned so long as information concerning the identity of the body prior to the discovery of the evidence contributing to E is available and is represented by the prior probabilities $\Pr(P), \Pr(Q), \Pr(R)$ and $\Pr(S)$. From Bayes' theorem (2.4)

$$\Pr(P \mid E) = \frac{\Pr(E \mid P) \Pr(P)}{\Pr(E)}$$

with similar results for Q, R, and S. From the law of total probability (1.11), since the events P, Q, R, and S are mutually exclusive and exhaustive

$$\begin{aligned}\Pr(E) = {} & \Pr(E \mid P)\Pr(P) + \Pr(E \mid Q)\Pr(Q) \\ & + \Pr(E \mid R)\Pr(R) + \Pr(E \mid S)\Pr(S).\end{aligned}$$

If the four prior probabilities are assumed mutually exclusive, exhaustive and equal,

$$\Pr(P) = \Pr(Q) = \Pr(R) = \Pr(S) = 1/4.$$

Hence,

$$\Pr(P \mid E) = \frac{\Pr(E \mid P)}{\begin{array}{c}\Pr(E \mid P) + \Pr(E \mid Q) + \Pr(E \mid R) \\ + \Pr(E \mid S)\end{array}}.$$

From Table 6.3

$$\Pr(P \mid E) = \frac{0.5000}{\begin{array}{c}0.5000 + 0.00135 \\ + 0.00183 + 0.00414\end{array}} = 0.9856,$$

the value given in the final row of the table. The posterior odds in favour of P versus Q, R, or S then equal $0.5000/(0.001\,35 + 0.001\,83 + 0.004\,14) \simeq 68$. This can be verified by determining $\Pr(P \mid E)/\Pr(\bar{P} \mid E) = 0.9856/0.0144 \simeq 68$ where $\bar{P}$ is the complement of P, which, for the moment, is taken to be Q, R, or S.

The value V of the evidence that equals $\Pr(E \mid P)/\Pr(E \mid \bar{P})$ may be determined, again if the prior

probabilities are available. From Bayes' theorem (2.4)

$$\begin{aligned}\Pr(E \mid \bar{P}) &= \frac{\Pr(\bar{P} \mid E)\Pr(E)}{\Pr(\bar{P})} \\ &= \frac{\{\Pr(Q \mid E) + \Pr(R \mid E) + \Pr(S \mid E)\}\Pr(E)}{\Pr(Q) + \Pr(R) + \Pr(S)} \\ &= \frac{\left\{\frac{\Pr(E|Q)\Pr(Q)}{\Pr(E)} + \frac{\Pr(E|R)\Pr(R)}{\Pr(E)} + \frac{\Pr(E|S)\Pr(S)}{\Pr(E)}\right\} \times \Pr(E)}{\Pr(Q) + \Pr(R) + \Pr(S)} \\ &= \frac{\Pr(E \mid Q)\Pr(Q) + \Pr(E \mid R)\Pr(R) + \Pr(E \mid S)\Pr(S)}{\Pr(Q) + \Pr(R) + \Pr(S)}. \qquad (6.8)\end{aligned}$$

If $\Pr(Q) = \Pr(R) = \Pr(S)$ then

$$\begin{aligned}\Pr(E \mid \bar{P}) &= \frac{\Pr(E \mid Q) + \Pr(E \mid R) + \Pr(E \mid S)}{3} \\ &= \frac{0.00732}{3} \\ &= 0.00244.\end{aligned}$$

Thus

$$\frac{\Pr(E \mid P)}{\Pr(E \mid \bar{P})} = \frac{0.5000}{0.00244} = 204.92.$$

It is about 205 times more probable to observe the evidence if the body is that of P than that of one of the other three women.

If $\Pr(P) = \Pr(Q) = \Pr(R) = \Pr(S) = 1/4$ then $\Pr(P)/\Pr(\bar{P}) = 1/3$ and the posterior odds in

favour of P equals

$$\frac{0.5000}{0.00244} \times \frac{1}{3} = \frac{0.5000}{0.00732} \simeq 68$$

as determined previously.

It is also possible to determine the value of the evidence in favour of P relative to one other woman, Q, say, in the usual way of comparing two propositions:

$$\frac{\Pr(E \mid P)}{\Pr(E \mid Q)} = \frac{0.5000}{0.00135} \simeq 370.$$

The evidence is about 370 times more probable to be obtained if the body is that of P rather than that of Q.

As well as age, the conditioning on background information I has, as usual, been omitted for clarity of exposition. However, there is information available regarding the length of time for which the women have been missing and their last known locations. This may be considered as background information I. As such it may be used to assign the prior probabilities that may then be written as $\Pr(P \mid I), \Pr(Q \mid I), \Pr(R \mid I)$, and $\Pr(S \mid I)$. Suppose $(P \mid I)$ is thought the most probable event and that the other three events are all equally improbable. Represent this as $\Pr(P \mid I) = 0.7$, $\Pr(Q \mid I) = \Pr(R \mid I) = \Pr(S \mid I) = 0.1$ (though this is not the only possible combination of probabilities which satisfy this criterion). The likelihood ratio $\Pr(E \mid P)/\Pr(E \mid \bar{P})$ is the same as before, 204.92, since it was determined assuming only that

$\Pr(Q) = \Pr(R) = \Pr(S)$, without specifying a particular value. The posterior odds alter, however. The prior odds are

$$\frac{\Pr(P \mid I)}{\Pr(\bar{P} \mid I)} = \frac{0.7}{0.3}.$$

The posterior odds then becomes

$$204.92 \times \frac{0.7}{0.3} \simeq 478$$

and $\Pr(P \mid E) = 0.998$.

The posterior probabilities in Table 6.3 have been calculated assuming $\Pr(P) = \Pr(Q) = \Pr(R) = \Pr(S) = 1/4$. If these probabilities are not equal, the posterior probabilities have to be calculated taking account of the relative values of the four individual probabilities. For example,

$$\Pr(P \mid E) = \frac{\Pr(E \mid P)\Pr(P)}{\begin{array}{l}\Pr(E \mid P)\Pr(P) + \Pr(E \mid Q)\Pr(Q) \\ + \Pr(E \mid R)\Pr(R) + \Pr(E \mid S)\Pr(S)\end{array}}.$$

The likelihood ratio also has to be calculated again if $\Pr(P), \Pr(Q), \Pr(R)$, and $\Pr(S)$ are not equal. From (6.8)

$$\frac{\Pr(E \mid P)}{\Pr(E \mid \bar{P})} = \frac{\Pr(E \mid P)\{\Pr(Q) + \Pr(R) + \Pr(S)\}}{\begin{array}{r}\Pr(E \mid Q)\Pr(Q) + \Pr(E \mid R)\Pr(R) \\ + \Pr(E \mid S)\Pr(S)\end{array}}. \tag{6.9}$$

These results for comparing four propositions may be generalised to any number, n, say, of

competing exclusive propositions. Let $H_1, \ldots, H_n$ be n exclusive propositions and let E be the evidence to be evaluated. Denote the probability of E under each of the n propositions by $\Pr(E \mid H_i)$, $(i = 1, \ldots, n)$. Let $\pi_i = \Pr(H_i)$, $(i = 1, \ldots, n)$ be the prior probabilities of the propositions, such that $\sum_{i=1}^{n} \pi_i = 1$. Then consider the value E for comparing H_1 with $(H_2, \ldots, H_n) = \bar{H}_1$, say.

$$\frac{\Pr(E \mid H_1)}{\Pr(E \mid \bar{H}_1)} = \frac{\Pr(E \mid H_1)(\sum_{i=2}^{n} \pi_i)}{\sum_{i=2}^{n} \Pr(E \mid H_i)\pi_i} \qquad (6.10)$$

from a straightforward extension of (6.9). Thus

$$\frac{\Pr(E \mid H_1)}{\Pr(E \mid \bar{H}_1)} = \frac{\Pr(E \mid H_1)(1 - \pi_1)}{\sum_{i=2}^{n} \Pr(E \mid H_i)\pi_i},$$

since $\sum_{i=2}^{n} \pi_i = 1 - \pi_1$, and

$$\begin{aligned} \Pr(H_1 \mid E) &= \frac{\Pr(E \mid H_1)\pi_1}{\Pr(E)} \\ &= \frac{\Pr(E \mid H_1)\pi_1}{\sum_{i=1}^{n} \Pr(E \mid H_i)\pi_i}. \end{aligned}$$

The posterior odds are best evaluated by writing $\Pr(\bar{H}_1 \mid E)$ as $1 - \Pr(H_1 \mid E)$ and then

$$\begin{aligned} \frac{\Pr(H_1 \mid E)}{\Pr(\bar{H}_1 \mid E)} &= \frac{\Pr(E \mid H_1)\pi_1 / \sum_{i=1}^{n} \Pr(E \mid H_i)\pi_i}{1 - \{\Pr(E \mid H_1)\pi_1 / \sum_{i=1}^{n} \Pr(E \mid H_i)\pi_i\}} \\ &= \frac{\Pr(E \mid H_1)\pi_1}{\sum_{i=2}^{n} \Pr(E \mid H_i)\pi_i}. \end{aligned}$$

The probability figures for non-sterilisation of women, and for women who are mothers and of a

certain age are based on information about white women in general. It is unlikely that the probabilities are the same amongst missing white women. However, these results are the best available and serve to illustrate the methodology.

There is a possibility not so far accounted for that the body may be that of a woman other than P, Q, R, or S. If such a possibility is to be considered, then it can be done by adding an additional proposition to the four under consideration and using the general results with $n = 5$. Information is then needed concerning the probability π_5 to be assigned to this proposition, remembering to adjust $\pi_1, \ldots, \pi_4$ appropriately so that $\pi_1 + \cdots + \pi_5 = 1$. Information is also needed concerning $\Pr(E \mid H_5)$. Some is available from Table 6.3 by taking relevant information from that pertaining to the other women. Thus Pr(mother| age unknown) = 0.583, from Q. However, Pr(not sterilised| mother unknown, age unknown) is unavailable. Probabilities are given in Table 6.3 for mothers of unknown age and for women of a certain age but about whom it is not known if they are mothers or not. A probability is not given for the probability that a woman would be sterilised when her age is unknown and it is not known whether she is a mother or not. The probabilities for E_3 (type A blood) and E_4 (other six phenotypes) would remain the same since the unknown woman has been identified from her remains as being white, and the appropriate probabilities have been given.

As previously noticed (see Section 6.1.6.1), it can be difficult to see how to apply directly (6.7) in a given scientific report because (6.7) directly relates to prior probabilities. The similar problem is faced with the full expression (6.10) for the value of the evidence. A solution to reduce the impact of prior probabilities has been proposed by Buckleton et al. (2006b) by simply considering an exclusive and exhaustive partition of $H_d = H_2, \ldots, H_n$ and by expanding it using the 'extension of conversation' rule (see Section 1.7.10). The value of the evidence requires just values for the prior probabilities $\Pr(H_i \mid H_d)$. The general expression is as follows:

$$\frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)} = \frac{\Pr(E \mid H_p, I)}{\sum_{i=2}^{n} \Pr(E \mid H_i, H_d, I) \Pr(H_i \mid H_d, I)}.$$

A general framework based on Bayes' factor for a number of situations is presented in Nordgaard et al. (2012b). Derivations for probabilistic graphical models are presented in Buckleton et al. (2006b), Biedermann et al. (2012d), and Taroni et al. (2014a).

6.1.7 A Note on Biological Traces

Consider likelihood ratio based evaluation in a common situation involving DNA where there is the profile E_r of the crime scene item and the profile E_c of the PoI. Let I represent the background information and let the propositions, for example, at the source level, be

H_p: the PoI is the source of the stain;

H_d: another person, unrelated to the PoI, is the source (i.e. the PoI is not the source of the stain).

Suppose that both profiles are, for example, of type A. From (5.2), the likelihood ratio can then be expressed as

$$\frac{\Pr(E_r = A \mid E_c = A, H_p, I)}{\Pr(E_r = A \mid E_c = A, H_d, I)}.$$

Assume that the DNA typing system is sufficiently reliable so that two samples from the same person will be found to have corresponding DNA profiles, and that there are no false negatives. Thus, the recovered item is expected to be of type A if it is known that the PoI is of type A, and if H_p is assumed true: $\Pr(E_r = A \mid E_c = A, H_p, I) = 1$.

It is widely assumed that the DNA profiles from two different people, such as the PoI and the donor of the stain when proposition H_d is true, are independent. Then $\Pr(E_r = A \mid E_c = A, H_d, I) = \Pr(E_r = A \mid I)$. In such a case only the so-called *profile probability* γ_A with which an unknown person would have the profile A is needed. In such a case γ_A is $2\gamma_i\gamma_j$ for a heterozygote stain/PoI and γ_i^2 for a homozygote stain/PoI, where γ_i and γ_j are the proportions of alleles *i* and *j* in a relevant population. This is a widely accepted simplification. In reality, however, the evidential value of a match between the profile of the recovered stain and the profile of the PoI needs to take into account

the fact that there is a person, the PoI, who has already been seen to have the profile of interest (here, type A). So, the probability of interest is $\Pr(E_r = A \mid E_c = A, H_d, I)$ and this can be quite different from $\Pr(E_r = A \mid I)$ (Weir, 2000b).

Observation of one gene in the (sub)population increases the chance of the observation of another of the same type. Hence, within a (sub)population, DNA profiles with matching allele types are more common than suggested by the independence assumption, even when two individuals are not directly related. The conditional genotype probability (also called *conditional match probability* or *random match probability*) incorporates the effect of population structure or other dependencies between individuals, such as that imposed by family relationships (Weir, 2000a).

The question thus is how many other individuals amongst the population of possible offenders might be expected to share the DNA profile of interest. The answer to this question is complicated by the phenomenon of genetic correlations due to shared ancestry (Balding, 1997). The calculation of profile probabilities using the so-called product rule (i.e. the multiplication of allele proportions within and across loci to obtain a probability for the complete multi-locus profile) is not sufficient when there are dependencies between different individuals involved in the case under examination, such as the PoI and the perpetrator (i.e. the actual source of the recovered stain) as assumed under the alternative proposition, H_d.

The more common source of dependency is a result of a membership of the same population and having similar evolutionary histories. The mere fact that populations are finite in size means that two people taken at random from a population have a non-zero chance of having relatively recent common ancestors. Thus, to disregard this correlation amongst alleles in the calculation of the value of the evidence is to exaggerate the strength of the evidence against the compared person (e.g. the PoI in a criminal case or the alleged father in civil paternity cases) even though it is not as important as the relatedness in the same population (Curran et al., 2003; Buckleton et al., 2006a).

A measure F_{ST} of inter-population variation in allele frequencies was introduced by Wright (1922). It can be considered as a measure of population structure. Extensive studies have been made of allele frequencies in many human populations to estimate values of F_{ST} (Balding et al., 1996; Foreman et al., 1997, 1998; Balding and Nichols, 1997; Lee et al., 2002) suggesting that it is prudent to use F_{ST} values from the large end of the currently observed range (Buckleton et al., 2016c). Values less than 0.01 are often used.

Evett and Weir (1998) discussed three so-called *F-statistics*, which provides a measure of relationship between a pair of alleles, and are relative to some background level of relationship. The notation used here is that used by Wright (1951, 1965). Evett and Weir (1998) use the notation of Cockerham (1969, 1973).

- The additional extent to which two alleles within one individual are related when compared with pairs of alleles in different individuals but within the same subpopulation is denoted F_{IS}.
- The extent of relatedness of alleles within an individual compared with alleles of different individuals in the whole population is denoted F_{IT}.
- The relationship between alleles of different individuals in one subpopulation when compared with pairs of alleles in different subpopulations, also known as the *coancestry coefficient*, is denoted F_{ST}.

Slight differences between the definitions of Wright (1951, 1965) and Cockerham (1969, 1973) are pointed out by Evett and Weir (1998). Wright defined his quantities for alleles identified by the gametes carrying them; Cockerham defined his statistics for alleles defined by the individuals carrying them. For random mating subpopulations, Evett and Weir (1998) comment that the distinction can be ignored. Discussion on the use of such statistics in forensic genetics is presented in Ayres and Overall (1999).

The following argument for deriving the probability of a correspondence is taken from Balding and Nichols (1994, pp. 138–139). Let γ_A and γ_B denote the population proportions of alleles A and B, respectively. Interpret the value of F_{ST} as the

probability that two alleles are identical through inheritance from a common ancestor in the same sub-population. With reasonable assumptions, two alleles drawn from the sub-population are identical through a common ancestor in the sub-population with probability F_{ST} and the ancestor is of type A with probability γ_A. If not identical by descent, two alleles are of type A with probability γ_A^2. Thus, the probability of drawing allele A in both of two random draws from the sub-population is

$$\Pr(A^2 \mid \gamma_A) = \gamma_A\{F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_A\}. \qquad (6.11)$$

The observation of one A allele in the subpopulation makes it likely that A is more common in the subpopulation than in the general population and hence $\Pr(A^2 \mid \gamma_A)$ is larger than the probability γ_A^2 of drawing two consecutive A alleles in the general population. The probability of drawing first an A followed by a B allele is

$$\Pr(AB \mid \gamma_A, \gamma_B) = \gamma_A\gamma_B(1 - F_{\mathrm{ST}}). \qquad (6.12)$$

In general, let $\Pr(A^rB^s)$ denote the probability that amongst $r + s$ alleles drawn randomly from the subpopulation, the first r are of type A and the following s are of type B then

$$\Pr(A^{r+1}B^s \mid \gamma_A, \gamma_B) = \Pr(A^rB^s \mid \gamma_A, \gamma_B) \times \frac{rF_{\mathrm{ST}} + \gamma_A(1 - F_{\mathrm{ST}})}{1 + (r + s - 1)F_{\mathrm{ST}}}. \qquad (6.13)$$

Special cases of (6.13) with $s = 0, r = 3$ and then $s = 0, r = 2$ gives

$$\Pr(A^4 \mid \gamma_A) = \Pr(A^2 \mid \gamma_A)\frac{\begin{array}{c}(2F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_A) \\ \times (3F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_A)\end{array}}{(1 + F_{\mathrm{ST}})(1 + 2F_{\mathrm{ST}})} \tag{6.14}$$

and

$$\Pr(A^2B^2 \mid \gamma_A, \gamma_B) = \Pr(AB \mid \gamma_A, \gamma_B) \times \frac{\begin{array}{c}(F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_A) \\ \times (F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_B)\end{array}}{(1 + F_{\mathrm{ST}})(1 + 2F_{\mathrm{ST}})}. \tag{6.15}$$

Assume that an innocent PoI, with profile $G_c = AB$ is drawn from the same subpopulation as the donor of the stain, who has profile $G_r = AB$ but that the two people are not closely related. Let A and B be the two observed alleles. Then

$$\begin{aligned}\Pr(G_r = AB \mid G_c = AB) &= \frac{\Pr(G_r = AB, G_c = AB)}{\Pr(G_c = AB)} \\ &= \Pr(A^2B^2)/\Pr(AB).\end{aligned} \tag{6.16}$$

Assume also that γ_A and γ_B are available only for a collection of subpopulations. Then from (6.15) and (6.16) the probability that the donor has a particular genotype given that the PoI has been found to have that type is

$$\Pr(G_r = AB \mid G_c = AB) = 2\frac{\begin{array}{c}\{F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_A\} \\ \times \{F_{\mathrm{ST}} + (1 - F_{\mathrm{ST}})\gamma_B\}\end{array}}{(1 + F_{\mathrm{ST}})(1 + 2F_{\mathrm{ST}})}, \tag{6.17}$$

where the factor 2 is to allow for the two ways of ordering the matching A and B alleles. This expression is known as the *conditional genotype probability*. Note that when $F_{ST} = 0$, the probability reduces to $2\gamma_A\gamma_B$, the basic result assuming Hardy–Weinberg equilibrium.

Similarly, the homozygote match probability may be obtained from (6.14), to give

$$\Pr(G_r = A^2 \mid G_c = A^2) = \frac{\{2F_{ST} + (1 - F_{ST})\gamma_A\}\{3F_{ST} + (1 - F_{ST})\gamma_A\}}{(1 + F_{ST})(1 + 2F_{ST})}.$$

These are the equations referred to in recommendation 4.2 of the 1996 NRC report (National Research Council, 1996). They allow the scientist to obtain profile probabilities for complete profiles. Single-locus probabilities from Balding and Nichols' (Balding and Nichols, 1994) formula are multiplied across loci. It should be emphasised that the results hold for two people in the same sub-population, but are an average over sub-populations. Allele proportions are an average over sub-populations and are not those in a particular sub-population. The last two equations allow population-wide allele proportions to be used for sub-populations for which F_{ST} applies. A simple derivation of Balding and Nichols' (Balding and Nichols, 1994) formula for heterozygotes and homozygotes is presented in Harbison and Buckleton (1998). Logical implications of applying the principles of population genetics to the evaluation of DNA evidence are presented in Triggs and Buckleton (2002).

It is possible to assess the effect of population substructure on forensic calculations. For heterozygotes between alleles with equal allele probabilities γ, the likelihood ratio for source level propositions is assessed as the reciprocal of the conditional genotype probabilities. Table 6.4 presents the likelihood ratios for various values of F_{ST}.

The effect of F_{ST} decreases as allele frequencies increase and is not substantial when $\gamma = 0.1$ even for F_{ST} as high as 0.01 (Weir, 1998). A discussion is presented in Taylor et al. (2014).

Therefore, it is important to distinguish between *profile probabilities* and *conditional genotype probabilities*. It is very helpful to use the term *profile probability* for the event of a single individual having a particular profile, in distinction to *(conditional) profile probability* for the event of a person having the profile when it is known that

Table 6.4 Effects of population structure as represented by F_{ST} on the likelihood ratio, the reciprocal of the conditional genotype probability (6.16) for heterozygotes between alleles with equal allele probabilities γ.

Allele probability	F_{ST}			
	0	0.001	0.01	0.05
$\gamma = 0.01$	5000	4152	1301	163
$\gamma = 0.05$	200	193	145	61
$\gamma = 0.1$	50	49	43	27

another person has the profile. The conditional profile probability, therefore, explicitly requires statements about two profiles. Profile probabilities are of some interest but are unlikely to be relevant in forensic calculations. It is of little consequence that the profile is rare in a given population, what is relevant is the rarity of the profile, given that one person (i.e. the perpetrator) has the profile. In other words, it is relevant to know the probability that the PoI would have the profile given that the perpetrator has the profile and that these are different people (Balding and Donnelly, 1995b). In practical cases, the pool of possible sources of a crime stain usually contains individuals with differing levels of ancestry shared with the PoI and therefore differing between-person correlations. The task is then to investigate the plausible range of conditional profile probabilities in a helpful and fair way. Further developments for mixed racial populations that avoid the approach of reporting separate values for each race is proposed by Triggs et al. (2000).

The distinction between profile and conditional profile probabilities is rarely made by practicing forensic scientists, and this is most likely because the two quantities have the same value in the simple case when 'product rule' calculations are assumed. If there is no relatedness in a large population, due to either immediate family membership or common evolutionary history, and there is completely random mating and population homogeneity, and an absence of

linkage, selection, mutation, and migration, then all the alleles in a DNA profile are independent. The profile probability and the conditional profile probability are both just the product of the allele probabilities, together with factors of two for each heterozygous locus. Thus, when $F_{ST} = 0$, the genotype probability reduces to γ_A^2, the basic result assuming Hardy–Weinberg equilibrium for homozygotes. The assumption that each of the alleles on a locus, one from the father and one from the mother, are independent of each other leads to an equilibrium distribution for the allelic proportions in a population. This is known as *Hardy–Weinberg equilibrium* or *random mating*. The allele population proportions are generally denoted with Latin letters, usually p. This differs from the convention used in the rest of the book of Greek letters denoting population proportions. However, the use of Roman letters, such as p (and q for $(1 - p)$), is so widespread that it is felt that the lesser confusion would be caused by following this practice. The genotype probability of AA is p_A^2, of Aa is $2pq$, and of aa is q^2, and $p^2 + 2pq + q^2 = (p + q)^2 = 1$. In general, let p_i and p_j be the population proportions of alleles A_i and A_j for $i, j = 1, \ldots, k$ where k is the number of alleles at the locus in question. The genotype population proportions (i.e. genotype probabilities) $\Pr_{ij}$ are obtained from the following equations assuming Hardy–Weinberg equilibrium:

$$\begin{aligned} \Pr_{ij} &= 2p_i p_j, \quad i \neq j, \\ &= p_i^2, \quad i = j. \end{aligned} \tag{6.18}$$

The previous discussion has assumed that the reference population contains no related individuals. Unrelated individuals have a very low probability of sharing the same profile but the probability increases for related individuals. In fact, relatives have the possibility of receiving the same genetic material from their common ancestors and therefore having the same DNA profile (Balding, 2000). So, the largest effect of dependencies between the DNA profiles of two individuals is when they are related. Other than for an identical twin, relationships such as brothers or fathers or cousins have very large effects on the likelihood ratio when their DNA profiles are not available. If a sibling or a close relative is amongst the possible contributors of the stain recovered at a crime scene, this should be reflected in the value of the likelihood ratio. Brothers, for example, have at least a 25% probability of sharing the same genotype at any locus. Consider, for example, the following two propositions:

- H_p: the PoI is the source of the crime scene item;
- H_d: a relative of the PoI left the crime item.

Let the evidence be the observation of alleles A_iA_j of the recovered and control item with allelic population proportions γ_i and γ_j. Assume that, if H_p is true, then the numerator of the likelihood ratio is 1, an assumption that will not necessarily be true but will serve to illustrate the point. If the donors of the recovered and control item are different individuals (i.e. H_d is true), and there is no familial relationship between them, then the denominator would be, under the simplifying

Hardy–Weinberg assumption, $2\gamma_i\gamma_j$ for $i \neq j$ and γ_i^2 for $i = j$. The effects of different familial relationships are given in Table 6.5 from Weir and Hill (1993) where numerical values are given assuming allelic population proportions of 0.1. See also Brookfield (1994), Ayres (2002), Ayres and Balding (2005), and Kaye (2016) for further examples and comments.

Sjerps and Kloosterman (1999) study cases in which the PoI's profile does not correspond to that of the crime stain, a result that may suggest that a close relative of the PoI might be found to correspond, in particular when the profiles share rare alleles.

Analyses using Bayesian networks have been presented in Dawid et al. (2002), Hepler and Weir (2008), Cowell et al. (2011), Cowell (2016), Taroni et al. (2014a), Mortera et al. (2016), Green and Mortera (2017), and Taylor et al. (2018a). A review of the use of probabilistic graphical models for DNA evidence assessment was presented in Biedermann and Taroni (2012).

Weir (2007) completes the previous conditional probabilities of Table 6.5 to account for the population structure parameter F_{ST} (see Table 6.6 where F_{ST} is denoted θ for clarity). Extension of this work has been published in Bright et al. (2013).

Conditional genotype probabilities are considered by Foreman and Evett (2001) for a variety of specified alternatives (i.e. possible sources of the stain other than the PoI) that correspond to individuals who exhibit different degrees

Table 6.5 Conditional genotype probability $\Pr(G_r \mid G_c, H_d, I)$ for the crime stain genotype (G_r), knowing the PoI's genotype (G_c), and assuming that a relative of the PoI is the donor of the crime stain.

PoI	Relative	$\Pr(G_r \mid G_c, H_d, I)$	Likelihood ratio, V
A_iA_j	Father or son	$(\gamma_i + \gamma_j)/2$	10
	Full brother	$(1 + \gamma_i + \gamma_j + 2\gamma_i\gamma_j)/4$	6.67
	Half-brother	$(\gamma_i + \gamma_j + 4\gamma_i\gamma_j)/4$	16.67
	Uncle or nephew	$(\gamma_i + \gamma_j + 4\gamma_i\gamma_j)/4$	16.67
	First cousin	$(\gamma_i + \gamma_j + 12\gamma_i\gamma_j)/8$	25
	Unrelated	$2\gamma_i\gamma_j$	50
A_iA_i	Father or son	γ_i	10
	Full brother	$(1 + \gamma_i)^2/4$	3.3
	Half-brother	$\gamma_i(1 + \gamma_i)/2$	18.2
	Uncle or nephew	$\gamma_i(1 + \gamma_i)/2$	18.2
	First cousin	$\gamma_i(1 + 3\gamma_i)/4$	30.8
	Unrelated	γ_i^2	100

Corresponding values for V assume allelic population proportions of 0.1.

Source: From Weir and Hill (1993). Reprinted with permissions of Elsevier.

of relatedness to the PoI and when there are fully corresponding profiles. The most common SGM-plus (10-locus STR profiling system) profile was determined using databases routinely used in forensic casework. Sampling error was taken into account using a size-bias correction (Curran et al., 2002; Curran and Buckleton, 2011). General

Table 6.6 Effects of family relatedness on conditional genotype probabilities, $\Pr(G_r \mid G_c, H_d, I)$.

PoI	Relationship	$\Pr(G_r \mid G_c, H_d, I)$
A_iA_j	Full sibs	$\frac{(1+\gamma_i+\gamma_j+2\gamma_i\gamma_j)+(5+3\gamma_i+3\gamma_j-4\gamma_i\gamma_j)\theta+2(4-2\gamma_i-2\gamma_j+\gamma_i\gamma_j)\theta^2}{4(1+\theta)(1+2\theta)}$
	Parent and child	$\frac{2\theta+(1-\theta)(\gamma_i+\gamma_j)}{2(1+\theta)}$
	Half sibs	$\frac{(\gamma_i+\gamma_j+4\gamma_i\gamma_j)+(2+5\gamma_i+5\gamma_j-8\gamma_i\gamma_j)\theta+(8-6\gamma_i-6\gamma_j+4\gamma_i\gamma_j)\theta^2}{4(1+\theta)(1+2\theta)}$
	First cousins	$\frac{(\gamma_i+\gamma_j+12\gamma_i\gamma_j)+(2+13\gamma_i+13\gamma_j-24\gamma_i\gamma_j)\theta+2(8-7\gamma_i-7\gamma_j+6\gamma_i\gamma_j)\theta^2}{8(1+\theta)(1+2\theta)}$
	Unrelated	$\frac{2[\theta+(1-\theta)\gamma_i][\theta+(1-\theta)\gamma_j]}{(1+\theta)(1+2\theta)}$
A_iA_i	Full sibs	$\frac{(1+\gamma_i)^2+(7+7\gamma_i-2\gamma_i^2)\theta+(16-9\gamma_i+\gamma_i^2)\theta^2}{4(1+\theta)(1+2\theta)}$
	Parent and child	$\frac{2\theta+(1-\theta)\gamma_i}{1+\theta}$
	Half sibs	$\frac{[2\theta+(1-\theta)\gamma_i][2+4\theta+(1-\theta)\gamma_i]}{2(1+\theta)(1+2\theta)}$
	First cousins	$\frac{[2\theta+(1-\theta)\gamma_i][1+11\theta+3(1-\theta)\gamma_i]}{4(1+\theta)(1+2\theta)}$
	Unrelated	$\frac{[2\theta+(1-\theta)\gamma_i][3\theta+(1-\theta)\gamma_i]}{(1+\theta)(1+2\theta)}$

Note the use of θ for F_{ST} for clarity.
Source: Adapted from Weir (2007).

conditional profile probability values range from 1 in 10 000 for a sibling relationship with the PoI to 1 in a billion for someone who is unrelated to the PoI. Values for other relationships such as parent/child and first cousins are given in Evett et al. (2000a). These values are recommended for use when reporting fully corresponding profiles. A general discussion of this topic is presented in Evett et al. (2000a). The same range of values is also reported by Hopwood et al. (2012) using 15-plex STR profiling systems.

Note that the use of values smaller than 10^{-9} for a conditional genotype probability is widely criticised (see, for example, Kaye (1993) and Curran (2010)).

Moreover, Hopwood et al. (2012) noticed:

> *Such values [values smaller than 10^{-9}] invoke independence assumptions to a scale of robustness that we cannot demonstrate empirically, given the size of available databases. […] In addition to the empirical evidence for the reliability of DNA evidence interpretation, we recognise also that such numbers are difficult to conceptualise and require unreasonable real life comparisons. (p. 188)*

Scientific (Lambert et al., 1995b; Weir, 2004; Curran et al., 2007, 2008; Buckleton et al., 2006a; Tvedebrink et al., 2012) and judicial (Koehler, 1997a; Saks and Koehler, 2008) literature has already underlined and discussed this practical aspect.

Hopwood et al. (2012) – crediting (Kaye, 2009c) – stated that:

> *A scientist quoting LRs higher than those currently presented, perhaps of the order of trillions or greater may serve only to detract from the real issues for the jury. (p. 188)*

Throughout this subsection, only a short overview of a specific issue associated with the evaluation of evidence from DNA profiles is given. Further and extended topics, such as DNA mixtures, evaluation of results for low template DNA, and Y-STR haplotypes, are given in specialist books, chapters on forensic DNA evidence, and scientific papers (e.g., Evett and Weir, 1998; Balding and Steele, 2015; Buckleton et al., 2016f,g; Taylor et al., 2018b; Bright et al., 2019), and in references therein.

The use of sophisticated software to deal with low template DNA and mixture stains has opened new topics for intensive discussion between scientists and lawyers on validation (e.g., Bright et al., 2015; Imwinkelried, 2016; Taylor et al., 2017b; Moretti et al., 2017; Kelly et al., 2018). See also *U.S. v. Gissantner* (2019) for comments on the support of software for evidence evaluation.

6.1.8 Additional Considerations on Source Level Propositions

Interest in the probabilistic evaluation of DNA profiling results has grown considerably during the past 20 years. Topics such as the consideration of error probabilities and the effect of database searches have been responsible for the increase

in interest in forensic inference amongst forensic scientists. Several aspects are developed below.

6.1.8.1 A Probabilistic Approach to Database Searching

When a scientific expert needs to assess the value of a DNA result, the manner in which the PoI was selected is crucial. The evaluation of DNA evidence for a PoI who was found through the use of a database has been a matter of scientific and judicial debate for some time, see, for example, Thompson and Ford (1989), Kaye (2001, 2009c), Balding (2002), Imwinkelried (2001). The debate was reopened when the German Stain Commission issued a recommendation supporting debated principles (Schneider et al., 2010) and there is a later contribution by Wixted et al. (2019).

The compilation of DNA databases seeks to enable investigative authorities to collect traces of unsolved criminal cases, as well as control material from convicted persons. Such stored information may help select PoIs in a way similar to the collection and storage of other types of forensic information, such as fingermarks, fingerprints, shoemarks, and shoeprints. Forensic databases are now well established across the world.

Confusion surrounding the interpretation of the outcome of a search in a database can arise because the probability of finding a correspondence increases as the database becomes larger. As explained by Robertson and Vignaux (1995a), the

confusion consists of claiming that the evidential value of a correspondence, say, of a DNA profile, when the PoI is selected through a search in a database, is affected by the number of comparisons that have been made. This leads to the erroneous conclusion that the larger the database, the weaker the evidence. This is one reason for believing that the evidential value of a corresponding DNA profile under such circumstances may be of little or no evidential value. However, evidence is still relevant if it is more (or less) probable to be observed if the PoI is the source than if an unknown person is the source. A correspondence between the PoI's DNA profile and that of a crime stain is certain, under some assumptions (e.g. absence of analytical error), when in fact the PoI is the source of the stain. However, corresponding DNA profiles are also certain, it is argued, when the result follows from a search of a database under the hypothesis that the PoI is not the source, because the PoI was chosen based on the fact that their DNA profile corresponds to the profile of the actual source of the crime stain. Therefore, the consequent likelihood ratio of 1 suggests that this evidence has no probative value: the evidence is as probable to arise if the PoI is the source as if an unknown person is (Thompson and Ford, 1989).

Confusion also arises because it is not clear whether the scientist is concerned with the probability of finding a match or with the increase that arises, through the discovery of the match, in the probability that it was the PoI who left the trace

(Robertson and Vignaux, 1995a). Analyses of the second point have been provided by Balding and Donnelly (1995b–1996), Dawid and Mortera (1996), and Evett and Weir (1998).

These analyses showed that the likelihood ratio is higher following a search in a database than in a case where the size of the potential criminal population is known and no sequential search has been performed. In fact, each person whose profile is found not to correspond with the DNA profile of the recovered trace is excluded. Therefore, the exclusion of these individuals from the size of the potential sources of the crime stain increases the probability that the individual, whose profile is found to correspond, is the source of the recovered stain. Although a database search is useful, it has to be emphasised that the strength of the overall case against the PoI can be much weaker than in the *probable cause setting*, defined by Balding and Donnelly (1996) as the setting in which the PoI has been identified on other grounds and subsequently subjected to DNA profiling. This is because of a lack of supporting evidence; no further incriminating evidence has been obtained (Balding and Donnelly (1996) and Donnelly and Friedman (1999)). Therefore, the discovery of a correspondence between the DNA profile of a crime stain and the profile of a person in a database does not mean that the perpetrator of the crime has been found.

The fact that the likelihood ratio for source level propositions is greater than the reciprocal of the

conditional profile probability may be justified through the arguments developed by Balding and Donnelly (1996) and Evett and Weir (1998).

Let H_p be the proposition that the PoI is the source of DNA found at the crime scene, H_d be the proposition that an unknown person is the source of the DNA found at the crime scene, and E be the evidence that the profile of the DNA found at the crime scene and the profile of the DNA of the PoI correspond. Then, the value, V, of the evidence is given by

$$V = \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)}.$$

A search has been made of a database which contains the DNA profiles of N named individuals. Exactly one of the profiles in the database corresponds to that of the DNA found at the crime scene and that individual becomes the PoI. Note that V does not depend on the probability that a search through the database would find a corresponding profile. The evidence is *not* that at least one (or exactly one) of the profiles of the individuals in the database corresponds to the profile of the crime stain. Other information does not affect the value of the evidence. Other information will be heard at any trial and will be accounted for there. To assess the value of the evidence, including the outcome of the search, let O denote the event that no other individual's profile in the database corresponds to the profile of the crime stain. Also, E may be separated into two components, E_r, the profile of

the crime stain and E_c, the profile of the PoI. Then, V can be written as

$$V = \frac{\Pr(E_r, E_c, O \mid H_p)}{\Pr(E_r, E_c, O \mid H_d)} = \frac{\Pr(E_r, E_c, \mid H_p, O)}{\Pr(E_r, E_c \mid H_d, O)} \frac{\Pr(O \mid H_p)}{\Pr(O \mid H_d)}. \quad (6.19)$$

In the probable cause setting, the value of the evidence is

$$\frac{\Pr(E_r, E_c, \mid H_p, O)}{\Pr(E_r, E_c \mid H_d, O)} \quad (6.20)$$

which is the first ratio in (6.19). Here the conditioning is extended to the information O. The numerator in (6.20) equals p, the profile probability. The fact that there has been a database search does not affect this probability. The denominator is given by the probability that two individuals chosen at random have corresponding profiles. Information O just increases one's confidence in the rarity of the profile. So,

$$\frac{\Pr(E_r, E_c, \mid H_p, O)}{\Pr(E_r, E_c \mid H_d, O)} = \frac{p}{p^2} = \frac{1}{p}.$$

It remains to be determined whether the second ratio, $\Pr(O \mid H_p)/\Pr(O \mid H_d)$ is smaller or greater than 1.

Consider $\bar{O}$. This is the event that at least one of the other individuals in the database matches the profile of the crime stain. If H_d is true there are two ways in which $\bar{O}$ may occur. One of the other individuals may be the source of the stain or none may

be the source but at least one happens by chance to match E_r. If H_p is true, then only the second of these is possible. Thus $\Pr(\bar{O} \mid H_p) < \Pr(\bar{O} \mid H_d)$. Hence $\Pr(O \mid H_p) > \Pr(O \mid H_d)$ and the second ratio in (6.19) is greater than unity. A more extended development that involves the general discriminating power of the profiling system for the assessment of the second ratio is proposed by Evett and Weir (1998). Their approach also shows that the second ratio is greater than unity. Thus, the value of the evidence when there has been a database search is greater than when the probable cause setting applies.

Balding and Donnelly (1996) and Evett and Weir (1998) argue that, although the difference in value is difficult to quantify in general, it seems likely that the database search value will be only slightly greater than the conventional likelihood ratio and that it is therefore convenient, and beneficial to the PoI, to calculate and report only the value without adjustment for the database search.

The aforementioned argument may be used to counter the following possible defence strategy. 'The profile probability for the DNA stain found at the scene of the crime is one in a million. The police database contains 10 000 profiles. The probability that, on searching the database, a match will be found is thus $10\,000 \times (1/1\,000\,000) = 1/100$. This figure, rather than one in a million, is the relevant probability to consider. This is not nearly small enough to be regarded as convincing evidence against the PoI'. From this point of view,

the effect of the database search is to weaken, very dramatically, the strength of the evidence against the PoI.

Consider a crime profile that has a profile probability of p. Consider a database with N unrelated individuals in it. The probability that the profile of an individual from the database does not correspond to the crime profile is $(1 - p)$ and, assuming independence, the probability that no profiles from the database corresponds to the crime profile is $(1 - p)^N$. Hence, the probability that at least one profile from the database corresponds to the crime scene profile, by chance alone, is $1 - (1 - p)^N$ (see Section 2.5.5), which, for small p such as occurs with DNA profiles, is approximately equal to Np, hence the figure $1/100$ in the previous paragraph. This result gives rise to the simple rule that, to determine the probability of a correspondence in a search of a database, one should take the match probability p and multiply it by the size N of the database.

This simple rule is concerned with the probability that there is at least one correspondence in the database. An extreme case can illustrate why the rule approximates to the answer for which it is designed and why it is not the right answer. Assume p, the profile probability, is extremely small and N is extremely large, such that Np is close to 1. It can be argued that the probability of finding at least one correspondence increases as N increases, and may even become close to 1 as N approaches the population of the world. Note that

the correct result is $1 - (1 - p)^N$ which will never be greater than 1. The simple rule, Np, cannot be used if it will give an answer greater than 1. As Np becomes larger, so the evidence reduces in value. However, the counter-argument is that the evidence becomes stronger as N becomes larger (Balding and Donnelly, 1996; Balding, 1997). This latter argument makes more sense, as it attaches greater value to the outcome of the search of a large database in which only one correspondence has been found when all other members of the database have been eliminated from the enquiry. The PoI so 'identified' is now one of a smaller overall population (smaller by the elimination of $(N - 1)$ members).

Balding and Donnelly (1996) make an interesting comment, in the light of subsequent discussion, see Stockmarr (1999), Dawid (2001), Devlin (2000), Evett et al. (2000b,c), Balding (2002), and Meester and Sjerps (2003). An alternative pair of propositions is that the source of the crime stain is or is not in the database. The probability of the evidence of exactly one correspondence in the database given the source is in the database is 1. The probability of the evidence of exactly one match in the database, given the source is not in the database, is $Np(1 - p)^{N-1}$. The likelihood ratio, assuming $(1 - p)^{N-1}$ to be 1, is then $1/Np$, which is the value given by the NRC (National Research Council, 1996). This result assumes that each of the individuals in the database is, without the DNA evidence, equally

likely to be the source. Note that it is possible for *Np* to be greater than 1, as has been pointed out by Donnelly and Friedman (1999). Such a result would imply that the evidence favours the PoI if a sufficiently large number of people are profiled. Thus one could have the rather strange situation where the more people that were profiled and found not to match the crime scene profile the more support there would be for the PoI's defence case.

The flaw in this argument is explained very well by Balding (2002). He notes that many statisticians will instinctively feel that a database search weakens the evidence against the PoI because one is conducting multiple comparisons. He comments that in a database search there is a 'crucial distinction', namely, that it is known in advance that exactly one proposition of the form '*X* is the source of the crime stain' is true. Consider two situations. In the first, there is evidence that a particular proposition is true. This would be the case if there had been no database search. In the second, there is evidence that a particular proposition is true and that many other propositions are false. This would be the case if there had been a database search. Put in this way, the evidence of a database search, in which no other correspondences were found strengthens the case against the PoI. The bigger the search which results in only a single correspondence, the more reason there is to be convinced that the observed correspondence is unique in the population.

A derivation is also offered by Berger et al. (2015). An extension for situations involving multiple correspondences is also presented in Robertson et al. (2016).

The derivation proposed by Berger et al. (2015) is as follows. Consider first the propositions of interest, say, H_p, the PoI is the person who donated the trace, and H_d, the trace was donated by some other (unrelated) person than the PoI. A likelihood ratio can be assigned by dividing the posterior to the prior odds, denoted $S_{H_p} \mid E$ and S_{H_p}, respectively, based on the expected number of alternative sources of the trace, say, S. The idea is therefore that of considering the factor that reduces the expected number of sources:

$$V = \frac{1/S_{H_p|E}}{1/S_{H_p}} = \frac{S_{H_p}}{S_{H_p|E}}.$$

Four quantities are of importance for the calculation:

N: the number of persons in the population;
n: the number of persons in the database;
m: the number of matches found in the database;
γ: the probability of finding the profile of interest in a relevant population.

At first, consider the classical scenario where a correspondence is reported between the recovered E_r and the control, E_c profiles.

$$V = \frac{S_{H_p}}{S_{H_p|E}} = \frac{N-1}{\gamma(N-1)} = \frac{1}{\gamma}.$$

As expressed by Berger et al. (2015, p. 157), 'the numerator is the number of other persons in the population $(N-1)$, and the denominator is the number of other persons in the population expected to match $\gamma(N-1)$'.

Consider now the situation involving a single match throughout a database search.

$$V = \frac{S_{H_p}}{S_{H_p|E}} = \frac{N-1}{\gamma(N-n)}.$$

Note that n persons have been excluded throughout the database search. So the denominator is reduced by n and the likelihood ratio becomes larger than in the first scenario. This solution is in agreement with previous solutions (e.g. Donnelly and Friedman 1999). Different scenarios are developed by the authors.

These insights have not always been well appreciated. In 2010, the German Stain Commission (Schneider et al., 2010) stated the following:

> *The German Stain Commission has developed recommendations on how to adequately take into account the probability of an adventitious match given the database size. Following these recommendations, the relevant match probability can be derived from the frequency of the DNA profile corrected by the actual number of persons in the database. (…) a statistical concept is described that allows to calculate either a match probability or a likelihood ratio without overestimating the weight of evidence following a database search. (p. 113)*

As noted in Taroni et al. (2011), elements of these recommendations have already been widely

discussed in both literature and practice, and found to be problematic. In a reply to Taroni et al. (2011), Fimmers et al. (2011) defended the Stain Commission's recommendations and reaffirmed the view that (i) the random match probability ought to be multiplied by a factor equal to the size of the database (which tends to reduce the value of the likelihood ratio), and that (ii) so-called data-dependent propositions (e.g. 'the PoI (some other person) is the source of the crime stain') should be avoided by choosing propositions of the kind 'the source of the crime stain is a person inside (outside) the database'. Based on a hypothetical example, the authors argued that this tends to reduce the probability of false convictions. However, the latter type of proposition is of little help because it suggests that the database is on trial; in reality, only the defendant is, hence the former set of propositions is appropriate.

A full discussion on the topic of DNA match statistics after a database search can be found in Walsh and Buckleton (2009) and Nordgaard et al. (2012b). Further probabilistic analyses of these issues are given in Biedermann et al. (2011c,e). An extension, using decision theory to analyse the issue of how to decide about the identification of the individual found as a result of a database search, is presented by Gittelson et al. (2012b).

Publication of Wixted et al. (2019) gave rise to a new debate on the management of a single-match DNA database search. The authors supported the

calculation of the posterior odds that the DNA belongs to the person who is the source of the single match by suggesting the evaluation of the prior odds based on a case-independent estimate of the size of the active criminal population as derived from database search statistics. A series of papers commented on and criticised the approach (see, Neumann and Ausdemore (2019), Meester and Slooten (2019a, 2019b) and Sjerps (2019) with rejoinders by Rouder et al. (2019) and Wixted and Rouder (2019)).

6.1.8.2 Search and Selection Effect (Double Counting Error)

Database searches are currently used by scientists in other areas of forensic science such as shoemarks and fingermarks. As previously, there may be a concern that the fact that the mark was retrieved as a result of a search in a database in some way weakens the evidence from the comparison with a shoe, for example. This is not the case (Evett et al., 1998). The likelihood ratio summarises all of the evidence that derives from the comparison. The fact that the mark (of the shoe) was found from a search of a database is relevant to the forming of prior odds in favour of the proposition that the PoI is the offender for that case (it is assumed that the shoemark is relevant and that there is a clear connection between the shoe and the PoI). If there is no evidence other than the geography of the incident and arrest, then the prior odds would be small but in this

case they could presumably be increased by the evidence of the property that was found at the PoI's home (Evett et al., 1998).

Robertson et al. (2016) noted that to avoid the unjustified repeated use of items of evidence, they ought only be used once in relation to each issue. This means that a decision maker, at a given instance in the process, may use the evidence for a given purpose (e.g. arrest), and that later, another decision maker, at another instance in the process, may use the evidence for another purpose (e.g. determination of guilt at trial).

Robertson and Vignaux (1995b) gave the example of a person stopped because of bloodstains observed on their clothing. In such a case, one might be tempted to consider the bloodstains of lower value compared with a case in which the person was selected based on other information. However, as Robertson and Vignaux (1995b) argued, this is not correct, the probative value of the bloodstains is still given by the probability of this observation given the competing propositions of interest. A difference between the two situations, selection of the person because of their bloodstained clothing or because of other information, arises because there may be less information in the former than in the latter situation. Specifically, if the PoI has been selected based on other evidence, that other evidence may already have increased the prior odds. Such an increase of prior odds might not have occurred in the case where the only information is the

bloodstained clothing leading to arrest in the first place. This also illustrates the crucial point that reasoning about a single item of evidence must not be conflated with reasoning on the case as a whole, based on all the evidence. In an analogous way, Meester et al. (2006) emphasised the problem of the use of available data twice: first to identify the suspect and indeed to suspect that a crime has occurred and after that again in the computation of some measure for the value of the evidence.

6.1.8.3 The Island Problem

Consider an island on which there are $N + 2$ people, one of whom is murdered. Of the remaining $N + 1$ people, N are innocent and one is the offender. Label these from 0 to N. There is trace evidence, such as DNA, thought to be left by the offender. The conditional profile probability for this item of evidence in the relevant population is γ. The posterior probability that a person with this profile is guilty is

$$1/(1 + N\gamma).$$

The arguments in this section are based on those expressed in Balding and Donnelly (1995a,b) and Balding (2000). A discussion of extensions beyond the scope of this book are given in Balding and Donnelly (1995a) and Balding and Steele (2015). Other relevant publications are Balding (1995) and Dawid and Mortera (1996). Here, the development provides useful insight into ideas underlying

inferences for forensic identification. There are two propositions to be considered

H_p : the PoI is the offender;

H_d : an unknown person is the offender.

The background information is denoted I and this is assumed to be independent of the evidence. If H_d is true, one of the other members of the island population is the offender. Let C be the random variable denoting the criminal and s the identity of the PoI (s is one of $0, 1, \ldots, N$). Using Balding and Donnelly (1995a) notation that replaces H_p and H_d, the expression $C = s$ denotes that the PoI is the offender. The expression $C = x$ denotes that individual x is the offender.

The evidence, E, is the DNA profile observed for the crime stain. A PoI s has been apprehended and observed to have the DNA profile. Proposition H_d is that an unknown person, different from the PoI, is the offender, and this can be denoted $C \neq s$.

The probability of the PoI being the offender, $C = s$, assuming independence between E and I may be written as

$$\Pr(C = s \mid E, I) = \frac{\Pr(E \mid C = s)\Pr(C = s \mid I)}{\Pr(E \mid C = s)\Pr(C = s \mid I) + \sum_{x \neq s} \Pr(E \mid C = x)\Pr(C = x \mid I)}. \tag{6.21}$$

Let $V_s(x)$ denote the likelihood ratio for x versus s (note that propositions H_p and H_d are switched),

$$V_s(x) = \frac{\Pr(E \mid C = x, I)}{\Pr(E \mid C = s, I)}.$$

The notation with s as a subscript is indicative of the asymmetry of the context in that the evidence is being considered with relation to propositions H_p and H_d that s is or is not the criminal. Let $w_s(x)$ be defined by the ratio

$$w_s(x) = \frac{\Pr(C = x \mid I)}{\Pr(C = s \mid I)}.$$

This ratio is neither an evidential value, since I is the conditioning, nor an odds since the propositions $C = s$ and $C = x$ are not complementary. It is interesting to consider the relationships between different situations and the values of $w_s(x)$, as discussed in Balding (2000). If the case against s rests primarily on DNA evidence there may be many x for which $w_s(x) \simeq 1$. Balding and Donnelly (1995b) noticed that

> *The perceived strength of DNA evidence has led to cases in which there is little or no evidence against the defendant other than the DNA evidence. Even if close relatives are unequivocally excluded, there may be many unrelated individuals who, if not for the DNA evidence, would be in much the same situation as the defendant. In other words, there may be many* i *such that* $\Pr(C = i)/\Pr(C = s)$ *is close to unity. (p. 11744)*

For most sexual or violent crimes $w_s(x) \simeq 0$ when x refers to women, children, and invalids. If there is strong alibi evidence or the victim has not been able to identify s then $w_s(x) \gg 1$. With this notation, (6.21) may be written as

$$\Pr(C = s \mid E, I) = \frac{1}{1 + \sum_{x \neq s} V_s(x) w_s(x)},$$

a result that has already been presented in (6.7) with a different notation. An example of this result has been given in Section 6.1.6.2.

The island problem is of some interest for the study of related questions, such as the evaluation of results of database searches (Section 6.1.8.1). As shown by Kaye (2009b), the logic of dealing with database search results can be illustrated through an extension of the island problem. Assume that the variable N represents the size of the total population. Suppose that the DNA profiles of the first $1, \ldots, n$ individuals, where the index 1 is that of the PoI, are in a database. The individuals $(n+1), \ldots, N$ are outside the database. Also one of the assumptions here is that the profile of the crime stain is compared with all n individuals in the database. This search of the database reveals that only the profile of the PoI corresponds to the profile of the crime stain. This correspondence is denoted by M_1. Besides, the database search has also revealed that the $2, \ldots, n$ individuals on the database other than the PoI do *not* correspond. The fact that a profile of an individual i, for $i = 2, \ldots, n$, does not correspond

to the crime stain is denoted here by X_i. One can thus write $X_2X_3\cdots X_n$ for the information that all entries of the database other than that of the PoI do not correspond. The two elements M_1 and $X_2X_3\cdots X_n$ need to be jointly evaluated. The totality of the findings may be written as $E_n = M_1X_2X_3\cdots X_n$.

Assume that the PoI will certainly correspond if they are the source of the crime stain, so $\Pr(M_1 \mid H_1) = 1$, where H_1 is that the PoI (individual 1 in the database) is the source of the crime stain. Let $\Pr(M_1 \mid H_i) = \gamma$ for $i = 2, \ldots, n$ where H_i is the proposition the profile of individual i corresponds to the crime stain but is not its source. Then π_1', the posterior probability for H_1 after considering the observation M_1, is

$$\Pr(H_1 \mid M_1) = \pi_1' = \frac{\pi_1}{\pi_1 + \gamma(1-\pi_1)} \qquad (6.22)$$

since $\Pr(M_1 \mid H_i) = \gamma$ for all i and thus

$$\begin{aligned}\sum_{i=2}^{N} \Pr(M_1 \mid H_i)\Pr(H_i) &= \gamma \sum_{i=2}^{N} \Pr(H_i) \\ &= \gamma(1-\pi).\end{aligned}$$

The size of the population is N. Consideration of the knowledge that there are $n-1$ individuals in a database (with $n < N$) leads to a minor refinement in the way in which the source level propositions H_i (for $i = 2, \ldots, n$) are formulated. In fact, they can now be framed as 'the profile of individual i in the database corresponds to the profile of the crime stain but is not its source'. A more

conceptual underpinning of these propositions H_i (for $i = 2, \ldots, n$) is that they refer to individuals who had their DNA profile compared with that of the crime stain. This is a difference with respect to the individuals $(n+1), \ldots, N$ whose profiles were not compared. On the whole, one can thus think of the population of size N as a splitting into n individuals as database members, and $N - n$ that are not. This splitting becomes apparent when writing the posterior probability for the observation E_n:

$$\Pr(H_1 \mid E_n) = \frac{\Pr(E_n \mid H_1)\Pr(H_1)}{\left(\begin{array}{l}\Pr(E_n \mid H_1)\Pr(H_1) \\ \quad + \sum_{i=2}^{n} \Pr(E_n \mid H_i)\Pr(H_i) \\ \quad\quad + \sum_{i=n+1}^{N} \Pr(E_n \mid H_i)\Pr(H_i)\end{array}\right)}. \tag{6.23}$$

This term can be shown to reduce to (Kaye, 2009b):

$$\begin{aligned}\Pr(H_1 \mid E_n) &= \pi_1' \\ &= \frac{\Pr(H_1)}{\Pr(H_1) + \gamma \sum_{i=n+1}^{N} \Pr(H_i)} \\ &= \frac{\pi_1}{\pi_1 + \gamma \sum_{i=n+1}^{N} \pi_i}.\end{aligned} \tag{6.24}$$

The logic of this result is that the second term in the denominator, $\gamma \sum_{i=n+1}^{N} \pi_i$, is smaller than $\gamma(1 - \pi_1)$ in (6.22). This latter expression involves a sum of prior probabilities for all members of the population (with no one except

the PoI being in the database), minus the PoI. In (6.24), the sum covers only those members of the population which are not in the database. Stated otherwise, the prior probabilities for the individuals in the database that are found to have profiles different from that of the crime stain are not relevant because of the multiplication with the zero likelihood, that is, $\Pr(E_n \mid H_i) = 0$, for $i = 2, \ldots, n$. Because of a smaller denominator, the posterior probability π_1' in (6.24) turns out to be greater than that in (6.22). The selection of a PoI in a database along with an exclusion of other database members by DNA profiling results thus provides more information against the corresponding PoI than if no database search had been conducted. These results can also be tracked in a Bayesian network as shown in Biedermann et al. (2011c–2012e) and Taroni et al. (2014a).

6.1.8.4 A Probabilistic Approach to Laboratory Error

Proper consideration of the role of the probability of error in the evaluation of DNA profiles is important. The need to consider the potential of errors in the assessment of forensic science evidence in general has been mentioned by many scholars in scientific and legal literature (Meier and Zabell, 1980; Gaudette, 1986, 1999; Koehler, 1996; Faigman et al., 2000; Saks and Koehler, 2005; Koehler, 2008, 2013, 2014, 2017a,b). Official reports on forensic science also insisted the importance of this topic. The Report to the US

President on 'Forensic science in criminal courts; ensuring scientific validity of feature-comparison methods', commonly known as the PCAST report (President's Council of Advisors on Science and Technology (PCAST), 2016), mentioned:

> *Without an appropriate estimate of accuracy, an examiner's statement that two samples are similar – or even indistinguishable – is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing - not training, personal experience nor professional practices - can substitute for adequate empirical demonstration of accuracy. (p. 46)*

When evaluating the strength of DNA evidence when propositions of interest relate to questions of source, two factors must thus be considered. One factor is the conditional profile probability. This value characterises the rarity of the DNA profile. The second factor is the probability of a false positive finding. A false positive occurs when a laboratory erroneously reports a correspondence in DNA profiles for two items that actually have different profiles. A false positive may occur due to error in the collection or handling of items, misinterpretation of test results, or incorrect reporting of test results (Thompson, 1995). Either a correspondence by chance or a false positive could cause a laboratory to report a DNA match between items from different people. Thus the conditional profile probability and the false positive probability should both be considered in order to make a fair evaluation of DNA evidence. Laboratory error rates, as determined, for example, in

proficiency testing, do not necessarily equate to the false positive probability in a particular case. The unique circumstances of each case may give rise to various different types of errors, but their probability of occurrence cannot be reduced to a conjectured error rate. Nevertheless, data on the rate of various types of errors in proficiency testing can provide insight into the order of magnitude of values for a particular case (Thompson, 1997; Koehler, 1997b). When DNA evidence is presented in court, juries typically receive data relevant for the profile probability only (Kaye and Sensabaugh, 2000). Kaye (1993) clearly mention the importance of the quantification:

> *As with the likelihood ratio or other probabilities, however, the most reasonable response is to insist that no DNA results be admitted without information on the rate of false positives as determined by external proficiency testing. (p. 167)*

A further practical difficulty is the presentation of a logical framework that takes into account both the probability of corresponding DNA profiles and the probability of error. Various suggestions have been made (Robertson and Vignaux, 1995b; Balding and Donnelly, 1995b; Balding, 2000). As pointed out by Balding (2000), when a case is based primarily on DNA evidence, the prosecution ought to demonstrate that the relevant error probabilities are small. This is particularly important when the conditional profile probability is very small: the latter can be misleading unless the

relevant probability of error is also small. Further, evaluations should focus on types of errors, and their related probabilities, that could actually have occurred in the case at hand, not any type of unspecified event of error.

A formal framework for considering the role that error may play in determining the value of forensic DNA evidence in a particular case is presented in Thompson et al. (2003). According to this framework, even a small false positive probability can, in some circumstances, have a strong impact, so serious consideration has to be given to this probability assignment. Thus, the proper assignation of false positive probabilities can be crucial for assessing the value of DNA evidence. Consider two propositions,

H_p: a crime scene stain came from a PoI;

H_d: the crime scene stain came from an unknown person.

The evidence E is a report of a correspondence between the DNA profile of the PoI and the profile of a stain found on a crime scene. The conditional profile probability and the probability of a false positive both contribute to $\Pr(E \mid H_d)$. Let M denote a true correspondence. It is assumed that either

M: the PoI and the crime scene stain have corresponding DNA profiles *or*

$\bar{M}$: the PoI and the crime scene stain do not have corresponding DNA profiles.

From the law of total probability (1.12)

$$\begin{aligned}\Pr(E \mid H_p) = {} & \Pr(E \mid M, H_p)\Pr(M \mid H_p) \\ & + \Pr(E \mid \bar{M}, H_p)\Pr(\bar{M} \mid H_p)\end{aligned}$$

and

$$\begin{aligned}\Pr(E \mid H_d) = {} & \Pr(E \mid M, H_d)\Pr(M \mid H_d) \\ & + \Pr(E \mid \bar{M}, H_d)\Pr(\bar{M} \mid H_d).\end{aligned}$$

The value of the evidence is then

$$\frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} = \frac{\begin{array}{c}\Pr(E \mid M, H_p)\Pr(M \mid H_p) \\ + \Pr(E \mid \bar{M}, H_p)\Pr(\bar{M} \mid H_p)\end{array}}{\begin{array}{c}\Pr(E \mid M, H_d)\Pr(M \mid H_d) \\ + \Pr(E \mid \bar{M}, H_d)\Pr(\bar{M} \mid H_d)\end{array}}.$$

Assume that given M, the report E is independent of H_p and H_d: the probability that a correspondence will be reported if there really is a correspondence is not affected by whether the correspondence is coincidental. Consequently, $\Pr(E \mid M, H_p) = \Pr(E \mid M, H_d) = \Pr(E \mid M)$. Consider further that the PoI and the crime scene stain will necessarily have corresponding DNA profiles if the PoI is the source of the stain, so $\Pr(M \mid H_p) = 1$ and $\Pr(\bar{M} \mid H_p) = 0$. Finally, because $\bar{M}$ can only arise under H_d, $\Pr(E \mid \bar{M}, H_d)$ can be simplified to $\Pr(E \mid \bar{M})$. Thus, the likelihood ratio becomes

$$\frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} = \frac{\Pr(E \mid M)}{\begin{array}{c}\Pr(E \mid M)\Pr(M \mid H_d) \\ + \Pr(E \mid \bar{M})\Pr(\bar{M} \mid H_d)\end{array}}.$$

In this expanded version of the likelihood ratio, the term $\Pr(E \mid M)$ is the probability that the analyst will report a correspondence if the PoI and the crime scene stain have corresponding DNA profiles and it is assumed to be 1.

The term $\Pr(M \mid H_d)$ is the probability of a coincidental correspondence of DNA profiles. For cases involving single-source items, $\Pr(M \mid H_d)$ is the (conditional) profile probability, denoted γ, and $\Pr(\bar{M} \mid H_d)$ is the complement of γ. The term $\Pr(E \mid \bar{M})$ is the false positive probability denoted ϵ. Thus

$$\frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} = \frac{1}{\gamma + \{\epsilon(1-\gamma)\}}.$$

The influence of variations in the prior odds in favour of H_p, γ, and ϵ on the posterior odds that the PoI was the source of the crime scene stain is shown in Table 6.7.

The prior odds presented in Table 6.7 are designed to correspond to two distinct case types that vary in how strongly the PoI is implicated as the source of the specimen by evidence other than the DNA results. Prior odds of $2:1$ describe a case in which the other evidence is fairly strong but not sufficient in itself for propositions of common source. It has been reported that DNA analysis results lead to the exclusion of approximately $1/3$ of PoI's in sexual assault cases. Hence, prior odds of $2:1$ might describe a typical sexual assault case submitted for DNA testing.

Table 6.7 Posterior odds that a PoI is the source of a crime scene stain that reportedly has a corresponding DNA profile, as a function of prior odds, random match probability, and false positive probability.

Prior odds	Random match probability	Probability of a false positive	Posterior odds
2 : 1	10^{-9}	0	2 000 000 000
2 : 1	10^{-9}	0.0001	20 000
2 : 1	10^{-6}	0	2 000 000
2 : 1	10^{-6}	0.0001	19 802
1 : 1000	10^{-9}	0	1 000 000
1 : 1000	10^{-9}	0.0001	10
1 : 1000	10^{-6}	0	1 000
1 : 1000	10^{-6}	0.0001	9.9

Source: Modified from Thompson et al. (2003).

Prior odds of 1 : 1 000 describe a case in which there is almost no evidence apart from the DNA result. The profile probabilities presented are chosen to represent two values that may plausibly arise in actual cases. Profile probabilities on the order of 1 in one billion (1 in 10^{-9}) are often reported when laboratories are able to find a correspondence between a reference item and a single source stain over 10 or more STR loci. Random match probabilities closer to 1 in one million (1 in 10^{-6}) are common when fewer loci are examined, for example, when the laboratory obtains only a partial profile for one of the examined items. The probability of a false positive in any particular case

will depend on a variety of factors. Some years ago, it has been suggested that the overall rate of false positives was between 1 in 100 and 1 in 1 000 (Thompson et al., 2003). New research on errors in a variety of forensic fields is now available (see, e.g. Langenburg et al. (2015), Song et al. (2018), Ribeiro et al. (2019), Martire et al. (2019), Murrie et al. (2019)).

Of course, for cases in which particular measures, such as repeated analyses, have been taken to reduce the possibility of the occurrence of an event of error, the false positive probability will be reduced. If two independent examinations comparing the same pair of items each had a false positive probability of 1 in 100, then the probability of a false positive on both examinations would be 1 in 10 000. A false positive probability of zero is also included for purposes of comparison. More results are available in Thompson et al. (2003). In the context of crime level propositions, Balding (2000) finds that the probability of a false negative is not relevant, at least to a first approximation.

It can happen that a PoI claims a reported correspondence, *RM*, between the recovered and control materials ($E_r = E_c$) arises because of a manipulation error made at the laboratory by the analyst (i.e. the analyst is suspected to have analysed an alternative material instead of E_c). This scenario can be analysed in a similar way to that arising from other forms of laboratory error and also taking account of a true match given a reported match. This analysis involves expanding

the conditional probabilities in the numerator and denominator of the likelihood ratio using the principles of the extension of the conversation (1.12).

A slightly different approach for combining the probability of laboratory error and the expression of the rarity of the corresponding DNA characteristics has been proposed in Buckleton et al. (2005). It is interesting to note that the approach in Buckleton et al. (2005) can be related to that of Thompson et al. (2003) by relaxing the assumption of no false negatives and by extending the view on the false positives (the probability of an error and the probability of a false positive correspondence given that an error has occurred). A formal comparison and the development of Bayesian networks are presented in Taroni et al. (2014a).

6.2 EXAMPLES FOR EVALUATION GIVEN ACTIVITY LEVEL PROPOSITIONS

In Sections 5.3.2.1 and 5.3.2.2, it has been shown that two derivations for the value of evidence can be obtained in cases where the findings are assessed given activity-level propositions. Let b_0 represent the probability of the presence by chance of 0 (groups of) stains/marks/traces, t_n and t'_n be the probabilities for transfer, persistence, and recovery of n stains/marks/traces from the person

of interest (or the alternative person, respectively), and $b_{g,\mathbf{m}}\gamma$ represent the probability of the chance occurrence of g groups of $\mathbf{m}$ stains/marks/traces recovered on the receptor ($b_{g,\mathbf{m}}$), with γ denoting the relevant population proportion of the observed characteristics and $\mathbf{m} = (m_1, \ldots, m_g)$ representing the group sizes. Probabilities t and b relate to what has been called *extrinsic evidence*, and proability γ to what has been called *intrinsic evidence* (Kind, 1994) (Section 5.3.2.1).

On one side, when the (transfer) material has potentially been left by the offender (i.e. transfer to the scene/victim), the likelihood can be expressed as follows (5.6):

$$V = \frac{b_0 t_n + b_{g,\mathbf{m}}\gamma t_0}{b_0 \gamma t'_n + b_{g,\mathbf{m}}\gamma t'_0}.$$

On the other side, when the material is found on the person of interest (potential transfer away from the scene), the likelihood ratio reduces to (5.7):

$$V = \frac{b_0 t_n + b_{g,\mathbf{m}}\gamma t_0}{b_{g,\mathbf{m}}\gamma}.$$

This equation can be simplified to

$$V = t_0 + \frac{b_0 t_n}{b_{g,\mathbf{m}}\gamma},$$

and approximated by $b_0 t_n / b_{g,\mathbf{m}}\gamma$. In what follows, examples of applications in fibre and glass domains are presented.

6.2.1 A Practical Approach to Fibres Evaluation

For the sake of illustration, consider the fibre cases introduced in Section 5.3.2.1, modified to versions (a) and (b) outlined later. Suppose that transfer probabilities have been assigned, following Curran et al. (1998b). It is assumed that the number of fibres transferred does not decrease as a function of time, because fibres are found on the back of a car seat, and that the performance of the recovery technique is high (between 90% and 95% of the fibres shed are recovered). The following results are obtained.

- (a): $\Pr(T_{10} \mid x, H_p) = t_{10} = 0.098$, $\Pr(T_0 \mid x, H_p) = t_0 = 0.005$, $\Pr(T_{10} \mid H_d) = t'_{10} = 0.0001$ and $\Pr(T_0 \mid H_d) = t'_0 = 0.0001$. Using conservative values for probabilities of background, say, $b_0 = 0.01$ and $b_{1,m} = 0.99$ (these assignments imply that it is very probable that a group of extraneous fibres will be found on the driver's seat, in that sense those values are conservative regarding the PoI). Note also that it is assumed that $b_0 + b_{1,m} \leq 1$ and that the occurrence γ for the fibre characteristics is set equal to 0.01. The following likelihood ratio is obtained:

$$V = \frac{b_0 t_{10} + b_{1,m} \gamma t_0}{b_0 \gamma t'_{10} + b_{1,m} \gamma t'_0}$$

$$= \frac{0.01 \times 0.098 + 0.99 \times 0.01 \times 0.005}{0.01 \times 0.01 \times 0.0001 + 0.99 \times 0.01 \times 0.0001}$$

$$\approx 1030.$$

This value strongly supports proposition H_p.

- (b): $\Pr(T_{10} \mid x, H_p) = t_{10} = 0.006$, $\Pr(T_0 \mid x, H_p) = t_0 = 0.0001$, $\Pr(T_{10} \mid H_d) = t'_{10} = 0.017$ and $\Pr(T_0 \mid H_d) = t'_0 = 0.021$. Using conservative values for probabilities of background, $b_0 = 0.01$ and $b_{1,m} = 0.99$, and a value of 0.01 for γ, the following likelihood ratio is obtained:

$$V = \frac{b_0 t_{10} + b_{1,m} \gamma t_0}{b_0 \gamma t'_{10} + b_{1,m} \gamma t'_0}$$

$$= \frac{0.01 \times 0.006 + 0.99 \times 0.01 \times 0.0001}{0.01 \times 0.01 \times 0.017 + 0.99 \times 0.01 \times 0.021}$$

$$= 0.29.$$

This likelihood ratio supports H_d, with a value of $0.29^{-1} = 3.4$.

These conclusions appear reasonable in view of the fact the number of recovered fibres, 10, is in agreement with transfer characteristics of the PoI's pullover and of the potential garments of the offender, respectively.

Various comments may be made following these examples. First of all, extraneous fibres have been collected on the upright part of the driver's seat. There is no information at all on potential fibres that may be on the bottom part of the seat. The

focus on multiple observations made at different positions on the seat, and the joint assessment of two or more items of evidence is an interesting topic, though potentially challenging using formal approaches.

Secondly, the probabilities for the background presence of fibres $b_{g,\mathbf{m}}$ relate to the occurrence by chance of g groups of $\mathbf{m}$ foreign fibres on the driver's seat. This kind of probability can be divided into two sets of probabilities, p_g, the probability by chance of g groups of fibres, and $\{s_{i,j_i}, i = 1, \ldots, g\}$ the probabilities of the ith group being of size j_i, as currently done in glass analysis, where j_i may take a positive integer value or be replaced by the letter l or s to denote that the group is *large* or *small*, respectively. Therefore, $b_{g,\mathbf{m}}$ can be replaced in formulae by $p_g \prod_{i=1}^{g} s_{i,j_i}$.

The next section (6.2.2) will present examples using such notation as proposed in the fibre pre-assessment example given in Section 5.5.3. For the assessment of glass fragments, it is assumed, as mentioned by Curran et al. (2000), that (i) there is no association between the number of groups found on surfaces of interest and the sizes of those groups, and (ii) there is no association between the occurrence of a given type of item with either the number of groups or the size of the group. These assumptions are questionable in the context of fibres evidence.

Thirdly, the trousers of the PoI have not been seized. It is of interest to look at these in relation with the driver's seat and consider the evidence potentially found on the two textile surfaces, the

trousers, and the bottom of the car seat. Suggestions on how to approach such a situation can be found in Taroni et al. (2014a) using Bayesian networks.

Fourthly, a pullover is supposed to have been in physical contact with the seat, so that it is expected that fibres from the pullover will be transferred to the seat. But note that fibres from the seat can also be transferred to the pullover. The scientist should be interested in a potential cross-transfer; see Section 5.3.2.5 for a suggestion on how to proceed to evaluate the two groups of findings in such a situation.

Finally, a grouping approach should be adopted for the recovered foreign fibres. Here the target group was defined as a set of items of material, which share the same forensic attributes. The scientist declares the presence of a group of fibres if there is sufficient characterisation in the shared features to relate these traces reasonably to a single source. However, this declaration amounts to a qualified opinion, and a formal approach on how grouping may be done should be explored. An example is presented in Triggs et al. (1997) and Curran et al. (2000) in the context of forensic glass analysis.

6.2.2 A Practical Approach to Glass Evaluation

Four situations are described and, for each, an expression for a likelihood ratio is derived.

Evett and Buckleton (1990) describe four situations involving the transfer of one or two groups of fragments of glass that may or may not have come from one or two windows that have been smashed during the commission of a crime.

The circumstances are as follows. One or two windows have been smashed with criminal intent. A PoI has been apprehended very soon after the crime and one or two groups of glass fragments have been found on their clothing. The two propositions of interest are

H_p: the PoI is the person who smashed the window(s) at the scene of the crime;

H_d: the PoI is not the person who smashed the window(s) at the scene of the crime; they have nothing to do with the incident.

Probabilities for various events need to be specified. These probabilities can be assigned by reference to an appropriate survey (as done historically by, for example, Pearson et al. (1971), Dabbs and Pearson (1970, 1972), Pounds and Smalldon (1978), Harrison et al. (1985), McQuillan and Edgar (1992), Lambert et al. (1995a), Allen and Scranage (1998), Allen et al. (1998a,b,c,d), Coulson et al. (2001a)) with care taken to ensure the relevance of the survey to the case in question and to the use of personal experience. The probabilities used here are those given by Evett and Buckleton (1990). The various events with their probabilities are:

- a person having no glass on their clothing by chance alone, probability $p_0 = 0.636$;
- a person having one group of fragments on their clothing by chance alone, probability $p_1 = 0.238$;
- a person having two groups of fragments on their clothing by chance alone, probability $p_2 = 0.087$;
- a person having more than two groups of fragments on their clothing by chance alone, probability $p_{2+} = 0{,}039$;
- a group of fragments found on members of the population being large, probability $s_l = 0.029$;
- during the commission of the crime, no glass being transferred, probability $t_0 = 0.2$;
- during the commission of the crime, a large group of fragments being transferred, retained, and found, probability $t_l = 0.6$.

The values are given for illustrative purposes and it is a simple matter to substitute other values where this is thought appropriate. Also, the definition of large is unspecified but again a suitable definition with an appropriate probability may be made for a particular case. If it is not felt possible to choose a particular value for a probability, then sensitivity analysis using a range of values may be tried. When a likelihood ratio remains relatively stable over the range of probability values investigated, this provides reassurance that a detailed elicitation of a given value is not crucial.

If a likelihood ratio does depend crucially on the choice of a probability, then careful thought is needed as to the usefulness of the method in the case under consideration.

For both windows, the values (γ_1, γ_2) for the occurrence of glass of the observed refractive indices on clothing is taken to be 3% in both cases, so that $\gamma_1 = \gamma_2 = 0.03$, where γ_1, γ_2 refer to the first and second windows, respectively. These values may be obtained from a histogram of refractive index measurements. In a more detailed approach, these values would be replaced by probability density estimates; see Section 7.5. Four cases can now be considered.

6.2.2.1 Case 1

One window is broken, one large group of fragments is found on the PoI, and it is found to have properties similar to those of the broken window.

The denominator, which is derived assuming that the PoI is innocent, is $p_1 s_l \gamma_1$, representing the product of the probability p_1, a person having one group of fragments on their clothing, the probability s_l, that such a group is large, and the value γ_1, reflecting the rarity of the fragments' analytical features in the relevant population.

The numerator is $p_0 t_l + p_1 s_l t_0 \gamma_1$. The first term accounts for the possibility that the PoI has had no glass on his clothing transferred by chance alone (p_0) and has had a large group of items transferred, retained, and found as a result of the commission of the crime (t_l). The probability that such a group

of fragments has the required properties is 1. The second term accounts for the probability that the PoI has had glass of the required properties present by chance alone ($p_1 s_l \gamma_1$) and no glass transferred as a result of the commission of the crime (t_0). The likelihood ratio for Case 1 thus takes the following form:[1]

$$V_1 = t_0 + \frac{p_0 t_l}{p_1 s_l \gamma_1}. \tag{6.25}$$

6.2.2.2 Case 2

One window is broken and two large groups of fragments (of size n and m) are found on the PoI. The analytical features of the first group correspond to those of the broken window. The second group does not correspond. The likelihood ratio for this case is

$$V_2 = \frac{t_0 2 p_2 s_n s_m \gamma_1 \gamma_2 + t_n p_1 s_m \gamma_2}{2 p_2 s_n s_m \gamma_1 \gamma_2},$$

where s_m and s_n are the probabilities that groups of size m and size n, respectively are found and recovered. Under H_p, the PoI smashed the window, two aspects are considered. First, no group of glass fragments has been transferred, has persisted, or was recovered, so the two groups are on the receptor for chance alone ($t_0 2 p_2 s_n s_m \gamma_1 \gamma_2$). Second, the group of n matching fragments has been

[1]The likelihood ratio can be, for sake of illustration, expressed in the following form: $t_0 + \frac{p_0 t_l}{p_1 s_l} \frac{1}{\gamma_1}$. Here, the value of γ_1 is set equal to 0.03. For a likelihood ratio assignment which takes continuous measurements into account, see Section 7.3.

transferred, has persisted, and was recovered, and the second group is present on the receptor for reasons unconnected with the criminal action ($t_n p_1 s_m \gamma_2$). The likelihood ratio can be simplified as follows:

$$V_2 = t_0 + \frac{p_1 t_n}{2 p_2 s_n \gamma_1}.$$

Note that the factor '2' appears in the denominator since there are two groups. As previously presented in Section 6.3.2, there are two possibilities to select two groups: you select first the group with n fragments and then the group of m fragments, or *vice-versa*. The variable accounting for the size and the probability of occurrence of the second group of fragments appears in both the numerator and denominator and cancel out. Hence, they do not appear in the final expression.

6.2.2.3 Case 3

Two windows are broken and one large group of fragments is found on the PoI, the properties of which correspond to one of the broken windows. In such a case, both propositions involve the two broken windows. If the PoI smashed the windows, the corresponding fragments are the result of transfer, persistence, and recovery, but the other group of fragments has not been transferred, persisted, or recovered. Otherwise both groups are present on the receptor by chance alone. If the PoI did not smash the two windows, the corresponding group of glass fragments is present by chance

alone. Given $t_{1,k}$ and $t_{2,h}$, the probabilities that k or h fragments have been transferred, persisted, and recovered, from the first and second windows, respectively, then the likelihood ratio becomes

$$V_3 = \frac{t_{1,0}t_{2,0}p_2s_k\gamma_1 + t_{1,k}t_{2,0}p_0}{p_1s_k\gamma_1}$$
$$= t_{1,0}t_{2,0} + \frac{t_{1,k}t_{2,0}p_0}{p_1s_k\gamma_1}.$$

The likelihood ratio can be reduced to

$$V_3 = t_0^2 + \frac{p_0t_0t_l}{p_1s_l\gamma_1},$$

when it has been assumed that the transfer probabilities $(t_{1,0}, t_{2,0})$ are the same for both windows and denoted t_0, and $t_{1,k}$ and s_k are denoted t_l and s_l with l denoting large.

6.2.2.4 Case 4

Two windows have been broken and two large groups of glass have been recovered on the surfaces of the clothing of a PoI. The analytical features of one group correspond to those of one broken window, and the features of the other group correspond to those of the other window. The proposition H_p specifies that the PoI is the person who smashed the two windows. Under H_d, the two recovered groups of glass fragments are present by chance. Under H_p, four possibilities should be taken into account:

(1) the two groups were transferred from the scene windows and the PoI has no glass on their clothing beforehand;

(2) one group of glass fragments came from scene window 1, and no glass was transferred from scene window 2. The PoI already had one group of glass fragments on their clothing;

(3) one group of glass fragments came from scene window 2, and no glass was transferred from scene window 1. The PoI already had one group of glass fragment on their clothing; and finally

(4) no glass fragments were transferred from the two scene windows, but the PoI already had two groups of glass fragments on their clothing beforehand.

The likelihood ratio becomes

$$V_4 = t_0^2 + \frac{p_0 t_l^2}{2p_2 s_l^2 \gamma_1 \gamma_2} + \frac{p_1 t_0 t_l}{2p_2 s_l \gamma_1} + \frac{p_1 t_0 t_l}{2p_2 s_l \gamma_2}.$$

Using the probability figures given earlier, it is easy to verify the second term is the dominant one.

Thus, in the four case examples considered so far, the following approximate results are obtained:

$$V_1 \simeq \frac{p_0 t_l}{p_1 s_l \gamma_1}, \tag{6.26}$$

$$V_2 \simeq \frac{p_1 t_l}{2 p_2 s_l \gamma_1}, \tag{6.27}$$

$$V_3 \simeq \frac{p_0 t_0 t_l}{p_1 s_l \gamma_1}, \tag{6.28}$$

$$V_4 \simeq \frac{p_0 t_l^2}{2 p_2 s_l^2 \gamma_1 \gamma_2}. \tag{6.29}$$

Using the probability values given earlier, with γ_1 and γ_2 equal to 0.03, Equations (6.26)–(6.29) give the following results:

$$V_1 \simeq 1\,843,\ V_2 \simeq 943,\ V_3 \simeq 368,$$
$$V_4 \simeq 1\,738\,000.$$

There are many imponderables, such as the specification of the transfer probabilities to be considered. Therefore, rather than in terms of their exact numerical values, a comparison of the orders of magnitude provides a useful qualitative assessment of the relative worth of these results. For example, consider a comparison of V_3 with V_4. The latter, V_4, is bigger than the former, V_3, by a factor of about 5000. The effect on the value of the evidence when two windows have been broken of discovering two groups of fragments, with similar properties to the broken windows, rather than just one, on the clothing of a PoI, is considerable. An approximate general formula is given by Curran et al. (2000) and more complex examples can be studied using probabilistic graphical models, such as Bayesian networks.

6.2.3 The Assignment of Probabilities for Transfer Events

The evaluation of transfer evidence given an activity level proposition requires assignments of probabilities for events of transfer. For example, in a case in which a window has been broken, consideration needs to be given to the probability of transfer, persistence, and recovery of glass fragments. These three factors may be referred to, generally, as the *transfer process*.

Assume that a crime is committed. Depending on the circumstances, it is expected by the investigators that there will have been transfer of evidence in both directions between the criminal and the crime scene. A PoI is apprehended and evidence is found on their person which is found to correspond, in some way, with material from the crime scene.

Transfer is an event that will depend on the nature of the contact between the criminal and the scene. For example, if a window has been broken to gain entry then the type of window and the distance from the window at which the criminal stood in order to break it will be factors in assessing the quantity of glass transferred.

Persistence is a phenomenon that will depend on the elapsed time between the commission of the crime and the apprehension of the PoI, their supposed activities during this time, and on the nature of the clothes that the PoI may be thought to have been wearing at the time of the crime.

Persistence will also depend on the nature of the contact, as for the transfer event. The persistence of glass fragments from a broken window may have features that are different from the persistence of blood following a prolonged assault.

Recovery will depend on the previous two factors, transfer and persistence. It will also depend on the quality of the resources, methods, and techniques available for the detection and collection of the evidence, including the examiner's skills in successfully operating the recovery procedure; see Samie et al. (2019) for a discussion focused on DNA evidence.

Note that the factors influencing the aforementioned three distinct phenomena may be explicitly expressed using Bayesian networks. Refer to Taroni et al. (2014a) for an example.

Modelling in the context of transfer probabilities constantly evolves and raises new issues of discussion, typically linked to the robustness and reliability of data supporting probability assignments. Samie (2019) explores such difficulties. A discussion is presented in Section 6.2.3.3.

Cook et al. (1993) provide an example of a case study involving fibres evidence. An assault was described in which a PoI answering a description given by the victim was arrested shortly after the assault was committed. Six fibres were found on the sweatshirt worn by the PoI, which were indistinguishable from those of a jumper worn by the victim. The probability of transfer, in this case, is the probability of more than one fibre of

the relevant type being transferred, persisting and being recovered from the PoI's clothing if the PoI committed the crime. Cook et al. (1993) note that, whilst it would be satisfying to consider the probability of exactly six fibres being found on the PoI's clothing, the imponderables of a particular case would not enable a probability distribution to be established with any degree of precision. Factors listed by Cook et al. (1993) that have to be considered include pressure and duration of contact, nature of the donor and recipient fibre surfaces, types of fibres involved, and elapsed time. These factors can form the basis of a discussion amongst investigators and probabilities for transfer of 0, 1 or more than one fibre can be determined by a consensus. These probabilities can be thought of as three positive numbers that add up to 1. Cook et al. (1993) suggested a graphical approach using a pie chart. Thus, a circle is to be split into three segments corresponding to the three probabilities for 0, 1 and more than one fibre being transferred, followed by persistence and recovery. The investigators can agree on the relative areas of the segments of the pie chart and the corresponding probabilities can be obtained. Such an approach will commend itself to those who find it easier to think visually rather than numerically.

Another example for the modelling of probabilities of transfer is given by Evett et al. (1995) in cases involving glass fragments as recovered material. The authors used a Poisson distribution

(Section A.2.6) for the number of glass fragments remaining at time t after the breaking of a window. Let X be the number of fragments remaining at time t, and let λ_t be the mean number of fragments remaining at time t, determined from experimental data on the persistence of glass fragments. Then

$$\Pr(X = x \mid \lambda_t) = \frac{e^{-\lambda_t}\lambda_t^x}{x!}, \quad x = 0, 1, \ldots.$$

Thus, given λ_t, the probabilities for different outcomes X are obtained. It is a matter of judgement by the expert, perhaps informed by experimental data, as to what value to choose for λ_t. However, as noted by Evett et al. (1995), the assumption of a Poisson distribution means that the variance of the distribution is also λ_t. This may give a value for the precision (thought of as the reciprocal of variance), which may not correspond to the expert's view. If this is the case, then, a different model will need to be assumed for the number of fragments transferred, followed by persistence and recovery, in a time interval of length t.

To assign probabilities for transfer, persistence, and recovery, a Bayesian point estimation procedure has been proposed in Biedermann et al. (2009b) for gunshot residues. Samie et al. (2016), in a DNA case involving a stabber and biological material transferred to a knife, assign probabilities for transfer based on the number of observations made in the experiments, using uniform prior counts. Point estimation, also called parameter

estimation, essentially refers to the process of using sample data to estimate the value of a population parameter. Sample surveys, which are a means for collecting sample data, commonly serve as a basis for parameter estimation. A Bayesian statistical model is specified with the choice of a prior distribution $\pi(\theta)$ that allows the scientist to express initial beliefs about the target parameter θ. Assuming a probability model for the data, all available information about the value of the parameter of interest θ, after observing the data, is contained in the posterior distribution $\pi(\theta \mid x)$. In other words, the posterior distribution encapsulates all that is known about θ. Examples of point estimation can be found in Taroni et al. (2010).

The relationships amongst the three relevant phenomena transfer, persistence, and recovery may be illustrated with a graphical model with nodes to represent these variables, and arrows joining the nodes to indicate relationships between the nodes thus joined. In addition, factors that contribute to the three variables of the transfer process may also be represented by nodes, with links to other nodes to indicate relationships as appropriate. An example of a graphical structure is presented in Taroni et al. (2014a).

It is possible to include probabilistic relationships in such a graph. For example, the probability of a transfer of trace material in a particular case may be dependent on several factors. This

dependency can then be represented with a conditional probability distribution, the conditioning being on the values of the factors on which the event of transfer is depending. The use of graphical models requires decisions to be made concerning not only the values of parameters in probability distributions but also the type of probability distributions themselves. For example, in the case of a broken window, the distance D from the window at which the criminal was standing when it was broken may be unknown. Because there is uncertainty associated with D, it may be modelled probabilistically. Suggestions have been presented by Curran et al. (1998b). Variable D has been modelled with a gamma distribution (A.3.5).[2] The distribution $\Pr(N = n)$ of the number N of fragments transferred is dependent on D and on other factors, such as those mentioned earlier. This distribution is not expressible as a formula from which probabilities may be simply obtained by the replacement of the values of the contributing factors. Instead, the distribution has to be derived empirically through a process known as simulation, which is beyond the scope of this book. Further details of the simulation process used for the modelling of the transfer of glass fragments are available in Curran et al. (2000). The authors commented (at p. 124) that the 'simulation process can be thought of as generating thousands of

[2]The software *tfer* (Forensic Glass Transfer Probabilities) developed by Curran is freely available, in R format, at https://cran.r-project.org/web/packages/tfer/ (last visited November 2019).

cases where the crime details are approximately the same and observing the number of fragments recovered'. For each of these 'thousands of cases', a value of n is obtained. A histogram of n is then derived and used as an approximation to the distribution $\Pr(N = n)$. From this approximation, an assignment for the probability at the particular value of N for the case in hand may be obtained.

Consider a case in which a PoI was apprehended about an hour or two after the commission of the crime. There is eye-witness evidence that the criminal was about one metre from the window when it was broken. The scientist expects that for a criminal standing one metre from the window in the circumstances of this crime that about 60 fragments of glass would have been transferred. The scientist expects about 80%–90% of any fragments transferred to the clothing of the criminal to be lost in the first hour and 50%–70% of the fragments remaining at the beginning of an hour to be lost in the subsequent hour. The scientist also expects to recover about 90%–95% of the fragments remaining on the clothing at the time of inspection.

The scientist inspects the clothing of the PoI and finds $N = 4$ fragments of glass. The simulation process of Curran et al. (1998b) indicates a probability $\Pr(N = 4)$ of 0.08 and $\Pr(N = 0)$ of 0.104. The values 0.08 and 0.104 are values for t_n in (6.25) solving Case 1 in Section 6.2.2, with $n = 4$ and $n = 0$, respectively. An example of the

application of such a methodology in fibres cases is presented in Champod and Taroni (2017).

As an example, consider another example using historical references for the factors of interest. The aim, at this point, is to describe the general methodology. Up-to-date data will substitute those used for the example. Let Γ be a DNA profile that has a conditional profile probability of approximately 0.01 amongst Caucasians in England. Assume that the distribution of DNA profiles amongst stains on clothing is approximately the distribution amongst the relevant population. This assumption is not necessarily correct (Gettinby, 1984), and there is further discussion of it below. Then $\gamma = 0.01$. A survey of men's clothing was conducted by Briggs (1978) from which it appears reasonable that $b_0 > 0.95$, $b_{1,1} < 0.05$. It is necessary to assign transfer probabilities t_0 and t_1 based on a study of the circumstances of the crime and, possibly, experimentation. Suppose $t_1 > 0.5$ (Evett, 1984). Then, irrespective of the value of t_0 (except that it has to be less than $(1 - t_1)$; i.e. less than 0.5 in this instance),

$$V > \frac{0.5 \times 0.95}{0.05 \times 0.01} = 950, \tag{6.30}$$

a value that represents strong support for the hypothesis that the PoI assaulted the victim, rather than they had nothing to do with the case. It is at least 950 times more probable to observe this evidence if the PoI assaulted the victim than if they did not.

Notice that the previous equation is considerably different from $1/\gamma$ (= 100 in the numerical example), a standard result if the findings are assessed given source-level propositions. The latter result would hold if $(t_1\ b_0/b_{1,1})$ were approximately 1, which may mean that unrealistic assumptions about the relative values of the probabilities of transfer would have to be made.

Various data have been published regarding studies carried out to investigate numerous aspects of transfer and persistence, and to offer values for probabilities. For glass particles, see references previously quoted in Section 6.2.2. A summary concerning transfer and persistence studies can be found in Curran et al. (2000).

Another example, by McDermott et al. (1999) concerns the transfer of paint fragments between vehicles, and foreign paint is found on a PoI's vehicle. Consider two propositions:

H_p: the PoI's vehicle and the injured party's vehicle were in contact;

H_d: the PoI's vehicle and the injured party's vehicle were not in contact.

A note on the use of the term 'contact' is in order. As much as the simple use of 'not' to frame the alternative proposition is unlikely to be particularly helpful to the court, it is also important to avoid the use of unclear words like 'contact' to describe the type of action specified by the propositions. There is a danger in using such a vague word. As emphasised by Evett et al. (2002b), the

statement that a PoI has been in recent contact, for example, with broken glass, could mean many things. There is a clear need to specify propositions concisely, considering the framework of circumstances, so as to ensure a transparent approach to the consideration of the evidence.

The value of the evidence in the example by McDermott et al. (1999) is approximately

$$V = \frac{t_n(1 - b_{g,\mathbf{m}})}{b_{g,\mathbf{m}}\gamma} \tag{6.31}$$

where

- t_n is the probability of paint transferring in the course of an automobile accident and consisting of one top coat layer;
- $b_{g,\mathbf{m}}$ is the probability that a random vehicle will have foreign paint on it, maybe in g groups of size $\mathbf{m}$;
- γ is the probability that the foreign paint would correspond to that of the injured party's vehicle (when paint transfer from the injured party's vehicle to the PoI's vehicle is considered).

McDermott et al. (1999) provided the following values. The value 0.8 is suggested for t_n on the basis of experience; paint consisting of at least a top layer is transferred in 80% of collisions investigated. A value for $b_{i,\mathbf{m}}$ of 0.094 is used, having been obtained from a survey of damage to vehicles in which 9.4% had foreign paint on the surface. A value of 0.127 is used for γ, this being

the proportion of vehicles with white solid paint. The value of the evidence is then

$$V = \frac{0.8 \times 0.906}{0.094 \times 0.127} = 61.$$

A similar argument applies when paint is transferred in the opposite direction, to the injured party's vehicle from the vehicle that caused the accident. It is important, however, not to approach this situation in terms of the simple reciprocal of the relevant population proportion of the analytical features of the recovered paint fragment, and the use of source level propositions of the following kind:

H_p: paint on the injured party's vehicle originated from the PoI's vehicle;

H_d: paint on the injured party's vehicle came from an unknown vehicle.

However, consider

H_p: the injured party's vehicle and the PoI's vehicle collided;

H_d: the injured party's vehicle collided with another vehicle.

With such propositions, the evaluation of the evidence has to take into account the possibility of transfer of paint from a source other than the PoI's vehicle. Assuming equal probabilities for background in the numerator and denominator of the likelihood ratio, the likelihood ratio V is

reduced to t_n/t'_n, an expression that emphasises the role of the transfer probabilities t_n and t'_n.

Notice again that the role of activity level propositions for the evaluation of transfer material can be illustrated through Bayesian networks. A review can be found in Taylor et al. (2018a, c).

6.2.3.1 Probabilities of Transfer in The Context of DNA

There is considerable ongoing research devoted to DNA evidence and transfer. This is due to the peculiar aspect of this type of evidence. Technical developments have made it possible to analyse very low amounts of DNA. This has many advantages, but the drawback of this technological progress is that interpretation of the results becomes increasingly complex: the number of mixed DNA profiles increased compared with single source DNA profiles and stochastic effects in the DNA profile, such as drop-in and drop-out,[3] are more often observed. Moreover, the relevance of low template DNA material regarding the alleged activities is not as straightforward as it was at a time when, for example, large quantities of blood

[3]Drop-in and drop-out are stochastic phenomena that appear in traces containing low quantities of DNA. Drop-out characterises the phenomena of an allele that fails to amplify (below the detection level). The phenomenon of drop-in refers to a false allele that is not reproducible. Note that drop-in should not be confused with contamination (or, pollution). The phenomenon of drop-in occurs independently for markers as opposed to contamination which affects the entire profile.

were recovered. The possibility of secondary and tertiary transfer is now becoming a pressing issue. This represents one of the core messages of the ENFSI (2015) guidelines for evaluative reporting.

It is regularly questioned, both during investigation and in court, how DNA traces were deposited, directly during the crime or innocently, for instance, by secondary transfer. Scientists cannot answer such questions directly. At best, using specialised knowledge, scientists can assign probabilities for findings *given* posited events of transfer or secondary transfer in different situations. The evaluation of evidence in such a context is described by Hicks et al. (2016) and Taylor et al. (2017a, 2019).

Several papers have been published on transfer-related topic. See, for example, Cale et al. (2016), Meakin et al. (2017), van Oorschot et al. (2017, 2019), and Burrill et al. (2019).

They present and consider data from trace DNA experiments to assess the type of DNA findings that may be obtained as a result of different types of transfer. Generally, the data show great variability in the results obtained, presumably as a consequence of the variety of factors that influence transfer, persistence, recovery, and analysis of trace DNA. A literature review of such experimental results is presented in Meakin and Jamieson (2013). Notice that scientists assign probabilities of transfer as one part (amongst others) of likelihood ratio based procedures for evaluating probative value (Evett and Weir, 1998,

pp. 35–39), but their main task is to address the question 'what is the probability to find a given quantity of material, if the PoI committed the alleged action'. Hence, scientists do not opine or report on the probability of transfer or what is often referred to loosely as 'the likely mode of transfer'. It is a completely different question to infer the action or mechanism of transfer based on a given quantity of recovered DNA. If done, using the experimental data only, this represents a case of a transposed conditional (or prosecutor's fallacy). In *R. v. D. Reed and T. Reed* (2009), it was noted:

> *It was common ground that cellular material could be transferred by direct contact with the plastic (primary transfer) by the person whose DNA had been so transferred or by secondary transfer to the plastic by another person transferring the cellular material of the person whose DNA had been found (or even tertiary transfer) […]. Although all of these were accepted by [the expert] to be possibilities, she expressed views that it was highly unlikely that the appellants had innocently touched the knives at some stage and that someone else had brought the knives to Peter Hoe's address. She also considered that it was unrealistic that each appellant had passed their DNA to someone else who then transferred it to the pieces of plastic which were found at Peter Hoe's address.*

Fonneløp et al. (2017) studied factors influencing transfer, such as the shedder status (high or low DNA shedders) of the involved persons, during stabbing activities. In the context of alleged sexual offences, evaluating the detection and non-detection of DNA corresponding to involved

persons is particularly challenging. To help with such cases, Jones et al. (2016) studied the rates of transfer of DNA in a variety of situations of staged non-intimate social contact as well as unprotected sexual intercourse.

Studies have also been devoted to different aspects relating to the mode of DNA deposition and transfer. For example, spermatozoa could regularly be obtained from bathwater (Page et al., 2014) and extraneous DNA can be found within an operational laboratory (Taylor et al., 2016b). Transfer and detection rates of a specific DNA source in the presence of background DNA sources have been studied (Lehmann et al., 2015). Issues in relation to who wore rather than touched a garment, in particular rates of DNA detection after wear and touch, have been studied by Breathnach et al. (2016). Transfer may also arise during the processing of items, in particular in laboratories. Margiotta et al. (2015) and Szkuta et al. (2015) have studied the role of gloves and laboratory examination tools in transfer. The occurrence of transfer beyond the laboratory, for example, by criminal investigators, has been studied by Fonneløp et al. (2016) and data on transfer and persistence of non-self DNA on hands over time can be found in Szkuta et al. (2018). The impact of the mode of DNA deposition and transfer on the value of evidence has been studied by Pun (2016).

In child sexual abuse cases, the alleged offender is often a close relative, such as a member of the same family. In such cases, it is relevant to

consider whether there is a possibility for the PoI's DNA to have been innocently deposited onto the victim's clothing, for example. One such possibility, investigated by Noël et al. (2016), is by the way of secondary transfer through a washing machine. Samie et al. (2016) studied the transfer of DNA when a person handles a knife. They also studied the potential of transfer of DNA from persons closely connected to the handler (Samie et al., 2020). In a wider perspective, Daly et al. (2012) investigated the transfer of DNA on different materials (glass, fabric, and wood). The recovery and analysis of transferred DNA, and the reporting of associated results, have also been studied and discussed in a comparative perspective, through an inter-laboratory exercise, with some variation found in the results obtained by the participants (Steensma et al., 2017). A discussion is presented in Section 6.2.3.3.

6.2.3.2 Transfer Probabilities and Micro-traces

Research on phenomena such as transfer has been pioneered by scientists working with traces other than DNA, in particular textile fibres, glass fragments, gunshot residues, and other transferable small particles. Research carried out is often a consequence of problems encountered during casework. An interesting example is the role of water in affecting fibre persistence. Indeed, many extraneous fibres present on a receptor surface can be lost as a result of exposure to rainfall or

immersion in water (Lepot and van den Driessche, 2015; Lepot et al., 2015). Weather and the skin of living persons were also found to play a major role. More generally, a body deposited outdoors is expected to lose the majority of transferred fibres within two days. However, no complete loss of fibres, in some studies focusing on time intervals up to twelve days, has been reported (Palmer and Burch, 2009; Palmer and Polwarth, 2011; Hong et al., 2014). Similar results have been obtained in a study on textile fibres on buried carcasses (De Battista et al., 2014). Yet other external influencing factors, such as washing machines, on transfer and the persistence phenomena of fibres have been studied by Watt et al. (2005). De Wael et al. (2010) have studied physical characteristics of textile fibres, such as sheddability. This study also proposes a practical method to assess the shedding potential of textile materials.

The occurrence of secondary transfer has been previously established for various evidence types. In the particular area of fibres, secondary transfer has been studied, for example, by French et al. (2012) and Palmer et al. (2017). Secondary transfer in chains involving both individuals and inert objects has been found to be possible, and can extend to tertiary and quaternary transfers. The population of textile fibres has also been studied in human hair. Palmer and Oliver (2004) found that the population of fibres in hair is comparable with other population studies involving other types of receptor surfaces. However, factors such as

season and geographical location may affect fibre populations in hair. Secondary transfer of fibres in hair, from a mask to pillow cases, has been studied by Palmer and Banks (2005), taking into account factors such as fibre type, hair style, time, and fibre persistence. A review of the literature is presented in Palmer (2016). General studies on fibres are also available (Wiggins et al., 2004; Coyle et al., 2012; Palmer, 2016).

Phenomena of transfer have also been studied extensively in relation to glass fragments, in the context of different modes of breaking windows (e.g. Hicks et al., 2005; Irwin, 2011; Cooper, 2013). A further common type of transfer material is paint. Moore et al. (2012) conducted a survey to investigate the background presence of paint fragments on items of clothing of persons suspected of involvement in crime, complementing previous research in this area by Pearson et al. (1971).

Other transfer materials that have been studied are pollen grain, powder, diatoms, and metal particulates, in particular regarding persistence on various types of target surfaces. Generally, it is observed that there is a decrease in the quantity of particles over time, comparable with what is known in existing literature about the persistence of fibres and glass fragements (Wiggins et al., 2002; Bull et al., 2006; French et al., 2012; Schield et al., 2016; Levin et al., 2017). Aspects of pollen distribution and persistence in a room have been studied in Morgan et al. (2014).

For the assessment of the probative value of gunshot residues (GSR), it is important to consider

the probability of finding GSR on a PoI by chance, that is when there was no recent direct exposure to a firearm discharge. Studies have thus been conducted to assess the presence by chance of GSR, and other particles resembling GSR, on members of the population at large (Grima et al., 2012; Lucas et al., 2016). Organic components of GSR, and in particular their persistence, have been studied by Arndt et al. (2012), whereas Hofstetter et al. (2017) focused on prevalence in different populations (general population and members of law enforcement personnel). Hannigan et al. (2015) studied the prevalence of GSR on clothing submitted to a laboratory in cases with no connection to firearms. The potential of GSR to undergo secondary and tertiary transfers has been studied by French and Morgan (2015). Conceptual aspects of evaluating GSR findings using likelihood ratios are discussed by Gauriot et al. (2013) and Gallidabino et al. (2015). Evaluation of evidence with regard to activity-related questions is more closely aligned to judicial and investigative aims than an assessment at source level propositions as already emphasised by the ENFSI (2015) guidelines and Maitre et al. (2017).

6.2.3.3 Additional Considerations on Activity Level Propositions and Transfer Probabilities

The evaluation of scientific results given activity level propositions represents an important topic for current forensic science practice. Illustrative examples for this can be found in connection with

DNA traces. When forensic scientists evaluate and report on the probative value of single donor DNA traces, it is common to rely on only one number expressing, in some sense, the rarity of the DNA profile in the population of interest (Section 5.3.1.1). The reason for this is that propositions of interest refer only to the source of the recovered trace material, such as 'The person of interest is the source of the crime stain' and 'An unknown person is the source of the crime stain'. Given the latter proposition, one is directed to think about the rarity of the DNA profile. Nowadays, however, DNA profiling technology is capable of producing results from very small quantities of trace material (e.g. non-visible staining). Such traces transfer very easily so that the issue of source is becoming less central, to the point that it is often not contested. As noted by Evett et al. (2002a), this has led to a shift from questions of the kind 'whose DNA is this?' to questions of the kind 'how did this DNA come to be there?'. This means that the primary need of recipients of expert information is assistance with evaluation when the competing propositions of interest refer to different posited activities (Taroni et al., 2013). This need is widely demonstrated in day-to-day forensic practice (Champod, 2013). Examples are complications encountered in cases such as *R v. Jama* (2009) (Vincent, 2010), where the PoI's DNA was found in trace material collected on a woman believed to have been sexually assaulted, or the Lukis Anderson case

(*People v. Howell* (2016)) (Kaplan, 2014) where DNA of the PoI supposedly made its way to the crime scene through the paramedics who had arrived at the victims' residence. A further example is *R. v. Weller* (2010) where alternative modes of transfer, such as the PoI touching the complainant's hair and vomit when putting her into bed, were raised as reasons for the DNA detected on the PoI's fingers, rather than placing his finger in the complainant's vagina. A report of the UK Forensic Science Regulator (Forensic Science Regulator, 2012) related to the case of Adam Scott who was also 'the innocent victim of avoidable contamination from an unrelated case that did contain his DNA'. (p. 13).

Notwithstanding the aforementioned, many forensic scientists remain reluctant to assess their results given propositions that relate to different activities. Some scientists consider evaluations beyond the issue of source as being overly speculative, because of the lack of relevant data and knowledge regarding phenomena and mechanisms of transfer, persistence, and background of DNA. Encouragement to assess findings given activity level propositions, as expressed in the ENFSI Guideline (ENFSI, 2015), highlights the need for rethinking current practice. Such proposals are sometimes viewed skeptically or are considered not feasible. Note that the ENFSI guideline clearly emphasised the following:

> *[I]t could be misleading to factually report the presence of two rare fibres on the victim that cannot be distinguished from the PoI's jacket when the circumstances of the case and the characteristics of the fabric suggest that a large number of fibres should have been found if the alleged activity has occurred. (p. 11)*

Generally, source-level propositions may be adequate in cases where there is no risk that the court will misinterpret them in the context of the alleged activities in the case. The ENFSI Guideline also presents practical examples to illustrate these understandings.

The extent to which available data can be used to support evaluation of DNA findings in the light of different activities sparks intensive discussion (Meakin and Jamieson, 2013, 2016; McKenna, 2013; Jamieson, 2016). Biedermann et al. (2016d) selected and discussed recurrent skeptical views, as well as some of the alternative solutions that have been suggested.

6.2.4 The Assignment of Probabilities for Background Traces

The general presence of transfer materials on various target surfaces (i.e. persons and their garments), commonly called background, is an important consideration in the evaluation of recovered trace material. Curran et al. (2000) summarised research back to the 1970s regarding glass found on different groups of persons (i.e. their clothing, head hair, etc.), connected

and unconnected with criminal activities. More recently, Daéid et al. (2009), for example, studied the presence of glass fragments on the clothing of law enforcement and forensic laboratory personnel. The population of foreign fibres on garments (or objects) has been studied, for example, by Marnane et al. (2006), and Cammarota et al. (2019), Lepot et al. (2017), Grieve et al. (2017), and Roux and Wiggins (2017). Moore et al. (2012) investigated the presence of paint flakes on the clothing of persons suspected of involvement in crime, whereas Reed et al. (2010) focused on polyurethane foam fragments on outer-garments. In the field of biological traces, the presence of saliva on underwear and bodily swabs is studied by Breathnach and Moore (2015) to help assess the probative value of saliva traces encountered in suspected sexual assault cases.

Consider a simple classic example based on the ABO blood group system. The aim of this example is to clarify the reasoning for the assignment of probabilities for background traces. In the previous section, it has been assumed that the distribution of blood groups among stains on clothing corresponds approximately to the distribution amongst members of the relevant population. This aspect has been questioned by Gettinby (1984). This is because the blood on a piece of clothing may have come from the wearer of the clothing and thus there is a bias in favour of the genotype of the wearer. The following argument is based on that in Gettinby (1984).

Consider a population of size N in which a proportion p have innocently acquired bloodstains on their clothing. Let p_o, p_a, p_b, and p_{ab} be the proportions with which blood groups O, A, B, and AB occur in the population such that $p_o + p_a + p_b + p_{ab} = 1$. Consider people of group O. Bloodstains detected on clothing may arise from several sources:

- by self-transfer (O type stains), with probability α, say,
- O stains from somewhere else, with probability β_0, say,
- stains of type A, B, or AB, necessarily from somewhere else, with probability γ_0, say,

such that

$$\alpha + \beta_0 + \gamma_0 = 1,$$
$$\beta_0 = (1 - \alpha)p_o. \qquad (6.32)$$

The proportion α is independent of the blood grouping of the individuals under consideration, unlike β_0 and γ_0. With an intuitively obvious notation, the following results for individuals of types A, B, and AB can be stated:

$$\alpha + \beta_a + \gamma_a = 1,$$
$$\beta_a = (1 - \alpha)p_a; \qquad (6.33)$$
$$\alpha + \beta_b + \gamma_b = 1,$$
$$\beta_b = (1 - \alpha)p_b; \qquad (6.34)$$
$$\alpha + \beta_{ab} + \gamma_{ab} = 1,$$
$$\beta_{ab} = (1 - \alpha)p_{ab}. \qquad (6.35)$$

Of those individuals who have bloodstains that have arisen from a source other than themselves (*non-self stains*), only a proportion γ will be distinguishable as such, where

$$\gamma = p_o\gamma_o + p_a\gamma_a + p_b\gamma_b + p_{ab}\gamma_{ab}.$$

For example, $p_o\gamma_o = \Pr(\text{type } A, B, \text{ or } AB \text{ stain found on clothing} \mid \text{person is of type } O) \times \Pr(\text{person is of type } O)$. Multiplication of pairs of Equations (6.32)–(6.35) by p_o, p_a, p_b, and p_{ab}, respectively, gives:

$$\begin{aligned} p_o\alpha + (1-\alpha)p_o^2 + p_o\gamma_o &= p_o, \\ p_a\alpha + (1-\alpha)p_a^2 + p_a\gamma_a &= p_a, \\ p_b\alpha + (1-\alpha)p_b^2 + p_b\gamma_b &= p_b, \\ p_{ab}\alpha + (1-\alpha)p_{ab}^2 + p_{ab}\gamma_{ab} &= p_{ab}, \end{aligned}$$

and summing gives

$$\alpha + (1-\alpha)(1-\delta) + \gamma = 1 \qquad (6.36)$$

where

$$\delta = 1 - p_o^2 - p_a^2 - p_b^2 - p_{ab}^2,$$

the discriminating power (Section 3.5) of the ABO system. From (6.36),

$$\alpha = 1 - \frac{\gamma}{\delta}.$$

Values of $\gamma = 0.182$ and $\delta = 0.602$ are used by Gettinby (1984) who cited Briggs (1978). From these a value of $\alpha \simeq 0.7$ is obtained for the probability of a bloodstain being acquired from

oneself, given that a bloodstain has been found on the clothing; i.e. approximately 70% of bloodstains on clothing are acquired by self-transfer.

Consider a person of blood group O. Denote by $C_O(O)$ the probability that this person innocently bears a bloodstain and that the bloodstain is of type O. Then

$$\begin{aligned} C_O(O) &= \Pr(\text{PoI has stain from self}) \\ &\quad + \Pr(\text{PoI has stain not from self but of type O}) \\ &= p\alpha + p(1-\alpha)p_o. \end{aligned}$$

With similar notation, for a person of blood group O to bear bloodstains of types A, B, or AB, the probabilities are

$$\begin{aligned} C_A(O) &= p(1-\alpha)p_a, \\ C_B(O) &= p(1-\alpha)p_b, \\ C_{AB}(O) &= p(1-\alpha)p_{ab}. \end{aligned}$$

The sum

$$C_O(O) + C_A(O) + C_B(O) + C_{AB}(O) = p,$$

is the probability of innocently acquiring a bloodstain. A value of $p = 0.369$ is given by Briggs (1978) and used by Gettinby (1984). Also, the distribution of blood groups amongst innocently acquired bloodstains on clothing of people of type O may be determined. For example, the probability a person of type O has a bloodstain of type O on his clothing, given it was acquired innocently is

$C_O(O)/p = \alpha + (1 - \alpha)p_o$. The distribution of blood groups, for people of type O, is thus

$$\Pr(\text{type } O \mid \text{innocent acquisition of a bloodstain}) = \alpha + (1 - \alpha)p_o,$$

$$\Pr(\text{type } A \mid \text{innocent acquisition of a bloodstain}) = (1 - \alpha)p_a,$$

$$\Pr(\text{type } B \mid \text{innocent acquisition of a bloodstain}) = (1 - \alpha)p_b,$$

$$\Pr(\text{type } AB \mid \text{innocent acquisition of a bloodstain}) = (1 - \alpha)p_{ab},$$

with similar results for people of types A, B, and AB. The comparison of this distribution with the general distribution is made in Table 6.8.

6.2.5 Presence of Material with Non-corresponding Features

In transfer evidence it may be that material present on the PoI or at the scene of the crime does not correspond to that found at the scene of the crime or on the PoI, respectively. For example, consider a case involving the potential transfer of fibres from the scene of a crime to the criminal. A PoI is found with fibres on his clothing which correspond, in some sense, to the fibres from the scene of the crime. However, the PoI also has fibres

Table 6.8 Distribution of blood groups of innocently acquired bloodstains on clothing of people of type O, compared with the distribution in the general population.

Blood group	O	A	B	AB	Total
Clothing of people of type O	$\alpha + (1-\alpha)p_o$	$(1-\alpha)p_a$	$(1-\alpha)p_b$	$(1-\alpha)p_{ab}$	1
General population	p_o	p_a	p_b	p_{ab}	1

of many different types on his clothing which do not correspond to those found at the scene (so-called foreign fibres).

A likelihood ratio for such a situation has been derived by Grieve and Dunlop (1992). It includes factors such as transfer probabilities, probabilities of foreign fibre types being found on the person, probabilities for the occurrence of the corresponding fibre types, and a factor to account for the number of corresponding fibres amongst the total number of fibres. The importance of this type of situation lies in the recognition that the number of items found, which do *not* correspond has to be considered when assessing the evidence as well as the number of items found which do correspond. A formal development has been presented in Section 5.5.3.

6.2.6 Absence of Evidence for Activity Level Propositions

6.2.6.1 A Question of Terminology

Terms like absence of evidence, negative evidence, or even missing evidence are often used as synonyms. A brief clarification note can be helpful. The notion *absence of evidence* can be considered as a generic term that encapsulates two aspects: on one side, the adjective *negative evidence* that specifies the non-occurrence of an expected event in cases where the scientist looks for a given item of evidence and they did not observe it (i.e. the

scientist did not observe the presence of a given DNA profile of interest on a receptor, but they observed another DNA profile or nothing at all). On the other side, *missing evidence* can be used to characterise the absence of information about the state of the expected evidence because the scientist did not report it (an example is presented in Section 6.3.5). The reason not to report does not play a fundamental role. Schum (1994) and Kadane and Schum (1996) agreed on the fact that the information is not acquired after a search of the evidence. But the use of the same phrase can also occur in situations in which the scientist did not even look for evidence. Therefore, by consequence, there is uncertainty about the presence (or not) of so-called positive evidence and about the existence of a situation involving negative evidence. A discussion can be found in Taroni et al. (2019).

6.2.6.2 Some Practical Examples

Based on a scenario presented by Hicks et al. (2016), consider the variation of some aspects and the development of two scenarios to see how the likelihood ratio, at activity level, can be quantified.

Scenario 1 A crime is committed and DNA is searched for on Mr A, a potential assailant, arrested 20 minutes after the alleged event. The prosecutor's view is that Mr A attacked Ms B who

spat several times on Mr A's face and T-shirt. Mr A said that he had never met Ms B. From the case information, it is known that Mr A was arrested in a bar, which he entered a few minutes after the incident and that he had not changed his T-shirt all day. The T-shirt of Mr A is searched for DNA and the forensic scientist noticed that there was only one single profile that corresponded to its owner (no extraneous DNA). Define the evidence E as that of no extraneous DNA. In order to evaluate this observation, the probabilities of these results under the competing propositions need to be assigned.

(1) If Ms B spat on Mr A (proposition H_p), and a single DNA profile that corresponds to the T-shirt wearer (Mr A) is recovered, this means that (i) there was no DNA transferred from Ms B when she spat on his T-shirt or that none was recovered (probability t_0) and (ii) there was no background DNA (extraneous DNA from some unknown source) on Mr A's T-shirt (probability b_0).

(2) If Ms B never spat on Mr A (proposition H_d), then the single profile is due to absence of background DNA (probability b_0).

The likelihood ratio equals $(t_0 \times b_0)/b_0 = t_0$, a value smaller than 1, so that the finding supports the proposition H_d.

Scenario 2 Consider a variation of the previous scenario. It is not the victim who spits on the

assailant, but the assailant who spits on Ms B and her T-shirt. It is not contested that the assailant spat on Ms B, but Mr A says he has nothing to do with the incident. DNA corresponding to Ms B only (the wearer of the T-shirt in this case) is recovered. Let E denote the evidence of no extraneous DNA. As in the previous scenario, probabilities of the findings should be quantified under competing propositions.

(1) If Mr A spat on Ms B (proposition H_p), and a single DNA profile that corresponds to the T-shirt wearer (Ms B) is recovered, no DNA was transferred from Mr A when he spat on her T-shirt (probability t_0) and no background on Ms B's T-shirt was present (probability b_0).

(2) If an unknown person spat on Ms B (proposition H_d), and no extraneous DNA is found (a single profile that corresponds to Ms B is recovered), no DNA was transferred from the unknown aggressor when he spat on her T-shirt (probability t'_0 and no background was present on Ms B's T-shirt (probability b_0).

The likelihood ratio becomes $(t_0 \times b_0)/(t'_0 \times b_0) = 1$ assuming that transfer probabilities are the same for the PoI and the unknown offender. The evidence is neutral and supports neither propositions.

More examples are presented and discussed in Taroni et al. (2019) in support of the analysis proposed by Thompson and Scurich (2018).

6.3 EXAMPLES FOR EVALUATION GIVEN OFFENCE LEVEL PROPOSITIONS

6.3.1 One Stain, *k* Offenders

Start by recalling the value of evidence expression (5.18), repeated here for convenience, when there are k offenders, one item of evidence (here a bloodstain), and the prosecution's proposition is that the PoI is one of the k offenders:

$$V = \frac{[r\{1+(k-1)\gamma\}/k] + \{\gamma'(1-r)\}}{\gamma r + \{p+(1-p)\gamma'\}(1-r)} = \frac{r\{1+(k-1)\gamma\} + k\gamma'(1-r)}{k[\gamma r + \{p+(1-p)\gamma'\}(1-r)]}. \quad (6.37)$$

Consider the case where it may be assumed that γ and γ', the occurrences of the feature characterising the recovered stain in a criminal population and in the general population, respectively, are (approximately) equal. This expresses the idea that there is no reason to suppose that the occurrence of a given genetic trait depends on the criminal origin of the donor (i.e. the donor is or is not one of the members of the group of k offenders). Notice that one may consider different population proportions regarding the analytical characteristic of the trace, for people with or without criminal background, typically if faced with findings such as textile fibres or shoemarks. Thus, the assumption $\gamma = \gamma'$ should be considered carefully.

Assume also that $p = 0$. This latter assumption holds if is considered impossible that the PoI may have left the stain at the crime scene for innocent reasons. Then

$$V = \frac{r\{1 + (k-1)\gamma\} + k\gamma(1-r)}{k\{\gamma r + \gamma(1-r)\}}$$
$$= \frac{r + (k-r)\gamma}{k\gamma}. \qquad (6.38)$$

If γ is so small that $r/k\gamma \gg 1$ then $V \simeq r/k\gamma$. If $r = 1, V \simeq 1/k\gamma$, see (6.4), the value of the evidence has been reduced by a factor corresponding to the number of offenders.

Consider another example. Assume $p \neq 0$ but that γ and γ' are (approximately) equal. Then

$$V = \frac{r + (k-r)\gamma}{k[p(1-r) + \gamma\{r + (1-p)(1-r)\}]}$$
$$= \frac{r + (k-r)\gamma}{k[p(1-r) + \gamma\{1 - p + pr\}]}.$$

A further interesting example of the use of such an approach relates to the evaluation of shoemarks, as given by Evett et al. (1998). The evaluation is analogous to that derived in Section 5.3.3.3, considering a single offender ($k = 1$). The propositions are, however, different. Consider a shoemarks case with E_c the evidence of a shoe-print from a PoI. The offence level propositions are

H_p: the PoI is the offender;

H_d: some unknown person is the offender.

The association propositions are

B: the shoemark was left by the offender;

$\bar{B}$: the shoemark was left by someone other than the offender.

The intermediate association propositions are

A: the shoemark was made by a particular shoe (X, say), owned by the PoI;

$\bar{A}$: the shoemark was made by some unknown shoe, which may or may not have been owned by the PoI.

Now, write $\Pr(A \mid H_p, B, E_c) = \Pr(A \mid H_p, B) = w$, the probability that the shoemark comes from the PoI's shoe X, given that the PoI was the offender and that the PoI left the shoemark. An illustration for the determination of w is given by Evett et al. (1998). The PoI was interviewed the day after the commission of the offence, he had ten pairs of shoes in his possession. Considering that the PoI has no preference for one of those pairs rather than for another, then w could be assigned as 0.1. Notice that in situations involving DNA evidence, for example, the value 1 can be retained for w. The reason for this is that a given person typically has, assuming the common DNA profiling systems, one type of DNA, though there may be exceptions to this, for example, with mt-DNA.

For shoemarks, a variant of (5.18) is available in which the conditional probability $\Pr(E_r \mid H_p, B, A, E_c)$, where E_r denotes the observed features in the mark, may be different from unity,

denoted p_{mrk} for short. The proportion of the relevant population of shoes that have the characteristics seen in the shoemark is γ. Consider, for sake of simplicity, γ and γ' to be equal. A more detailed analysis would treat these as different since the probability of seeing a particular shoemark may be dependent on the wearer. Evett et al. (1998) consider the rarity of the observed features to be of two parts, one relevant to the manufacturing features and one relevant to the acquired features. Also, assuming that the shoemark has not been left for an innocent reason enables p to be set equal to zero. Then $\Pr(E_r \mid H_p, B, E_c) = wp_{mrk} + \gamma(1-w)$ and $\Pr(E_r \mid H_p, \bar{B}, E_c) = \Pr(E_r \mid H_d, B, E_c) = \Pr(E_r \mid H_d, \bar{B}, E_c) = \gamma$. Then

$$\begin{aligned} V &= \frac{r\{wp_{mrk} + \gamma(1-w)\} + \gamma(1-r)}{\gamma r + \gamma(1-r)} \\ &= \frac{rwp_{mrk} - rw\gamma + \gamma}{\gamma} \\ &= (1-rw) + \frac{rw}{\gamma}p_{mrk}. \end{aligned} \tag{6.39}$$

Numerical examples are given in Evett et al. (1998).

An alternative way to express (6.39) is

$$V = rwV_mV_a + (1-rw),$$

where V_m and V_a represent the value of the evidence under source-level propositions for observable aspects of the shoemark related to manufacturing (m) and acquired (a) features,

respectively. So,

$$\begin{aligned}\Pr(E_r \mid E_c, A) &= \Pr(E_{ra}, E_{rm} \mid E_c, A) \\ &= \Pr(E_{ra} \mid E_{rm}, E_c, A)\Pr(E_{rm} \mid E_c, A).\end{aligned}$$

These conditional probabilities represent the denominator of V_a and V_m, respectively.

In this likelihood ratio development, it is important to devote some consideration to the probabilities r and p. The former, r, is the probability that the crime stain came from one of the offenders. Stoney (1991a) has referred to this probability as the *relevance* of the crime stain. The second probability, p, is the probability that the crime stain came from the PoI, given the PoI did not commit the crime and that the crime stain did not come from any of the offenders. Stated otherwise, this is the probability that the crime stain was left innocently by someone who is now a PoI.

Evett (1993a) suggested that determination of the probabilities such as those earlier may be the province of the court and that it is necessary to establish the conditions under which the scientific evidence can be of any guidance to the court. Evett suggested an examination of the sensitivity of V to values of p and of r. As an illustration, in a case involving a biological stain, the number of offenders k is taken to be 4 and the factors γ and γ' are set to 0.001. Then

$$V = \frac{r + 0.004}{4[p(1-r) + 0.001(1 - p + pr)]},$$

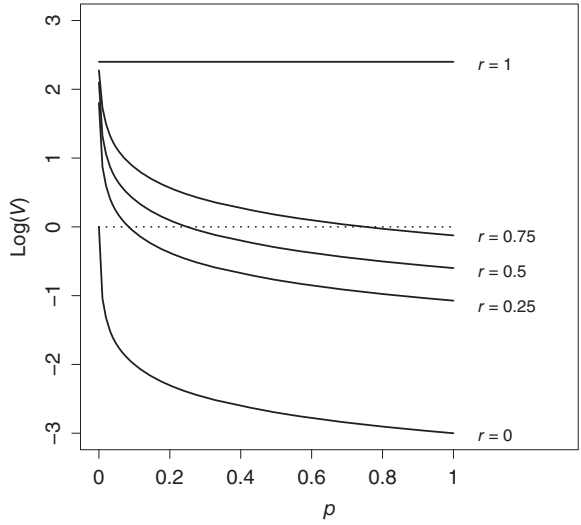

Figure 6.1 Variation in logarithm of the value of evidence with variation in the probability that the evidence had been left innocently and variation in the relevance.] Variation in the logarithm to base 10 of the likelihood ratio V of the evidence with p, the probability that the stain would have been left by the PoI even though they were innocent of the offence, for various values of r, the probability that the stain would have been left by one of the offenders. The number of offenders, k, equals 4 and the population proportion of the corresponding features in the mark, γ, is 0.001. Source: Adapted from Evett (1993a) with the inclusion of a curve for $r = 0$. The dotted line at $\log(V) = 0$ indicates where the evidence has the same probability under both propositions.

where $r + (k - r)\gamma$ has been approximated by $r + k\gamma$. The variation of V with r and p is shown in Figure 6.1.

The graph has been drawn with a logarithmic scale for V. This is plotted against p for $r = 1$, 0.75, 0.50, 0.25, and 0. It is useful

to consider individually the terms within the expression for V for the case in which there is one bloodstain of profile Γ and associated population proportion γ:

- The number of offenders, k: this factor is assumed to be well known. Note, however, that uncertainty may accompany k, similarly to DNA mixtures where the number of contributors to the mixture is not agreed by the parties. An approach to deal with uncertainty about the number of contributors is presented in Biedermann et al. (2011d).
- Relevance, r: the probability that the crime stain came from one of the offenders. Here, factors to be considered include location, abundance, and apparent freshness of the recovered blood.
- Innocent deposit, p: the probability that the crime stain came from the PoI, given that the PoI did not commit the crime and given that the crime stain did not come from any of the k offenders.

Values for the probabilities of relevance and innocent deposit may be retained by the scientist, in agreement with the parties, but some argue the values should be decided by the courts. In general, V decreases as r decreases or as p increases.

For $r = 1$, meaning that it is certain that the crime stain came from one of the offenders,

$$V = \frac{1.004}{0.004},$$

$$= 251,$$
$$\log_{10}(V) = 2.40.$$

For $r \neq 1$, V is very sensitive to p. If there is a non-zero probability that the crime stain did not come from one of the offenders, then the probability of innocent deposit has a considerable influence on V. For example, if $r = 0.25$, so that there is a small probability the crime stain came from one of the offenders, V becomes less than 1 for $p > 0.083 \simeq 1/12$. Thus, if $p > 1/12$ (and there is a small probability that the crime stain came from the PoI, conditional on everything else), then the evidence supports the proposition that some other person is one of the offenders rather than the proposition that the PoI is one of the offenders.

6.3.2 Two Stains, One Offender

The two-trace bloodstain problem of Section 6.3.2 has been modified by Stoney (1994) to the case where there are two bloodstains, of profile Γ_1 and Γ_2, respectively. There is only one offender, however, rather than two as in Section 6.3.2, who left one of the bloodstains, but it is not known which one. A PoI is found who is of group Γ_1. Relevance is applicable in this context since it provides a measure of the belief (probability) in the proposition that the stain at the crime scene that comes from the offender is the one that is of

the same group as the PoI. The two competing propositions to be considered are as follows:

H_p: the PoI is the offender;

H_d: an unknown person is the offender.

Let r be the probability that the stain (Γ_1) with corresponding profile was from the offender. It has been assumed that one of the stains is from the offender so there is a probability $(1 - r)$ that it is the other stain (Γ_2) that is from the offender. Suppose H_p is true. Then, using association propositions, the following may be considered:

$$\begin{aligned}
&\Pr(\text{PoI's profile and crime profile} \\
&\qquad \text{correspond in type } \Gamma_1 \\
&\mid \text{stain } \Gamma_1 \text{ was from the offender}) = 1. \\
&\Pr(\text{stain } \Gamma_1 \text{ was from the offender}) = r. \\
&\Pr(\text{PoI's profile and crime profile} \\
&\qquad \text{correspond in type } \Gamma_1 \\
&\mid \text{stain } \Gamma_2 \text{ was from the offender}) = 0. \\
&\Pr(\text{stain } \Gamma_2 \text{ was from the offender}) = (1 - r).
\end{aligned}$$

Thus $\Pr(\text{correspondence in type } \Gamma_1 \mid H_p) = r$.

Suppose next that H_d is true, so there is no need to develop the denominator of V using association propositions. The probability of a correspondence in Γ_1 is the probability that a randomly selected person is of profile Γ_1. This is the population

proportion γ_1. The likelihood ratio is then

$$V = r/\gamma_1.$$

If the two stains have equal probabilities of being left by the offender, so that $r = 1/2$, then the likelihood ratio equal $1/2\gamma_1$. This is numerically equivalent to the figure $1/2\gamma_1$ derived by Evett (1987b) and quoted in (6.3) for a problem with two stains, one left by each of the two offenders, and a single PoI whose profile corresponds to that of one of the stains and whose population proportion is γ_1. The derivation, however, is very different.

Stoney (1994) continued the development for the case where neither stain may be relevant but there is still a single offender. The PoI has profile Γ_1. Let there be the following probabilities

$$\begin{aligned}
&\Pr(\text{The stain of profile } \Gamma_1 \text{ is from} \\
&\qquad \text{the offender}) = r_1, \\
&\Pr(\text{The stain of profile } \Gamma_2 \text{ is from} \\
&\qquad \text{the offender}) = r_2, \\
&\Pr(\text{Neither stain is from} \\
&\qquad \text{the offender}) = 1 - r_1 - r_2.
\end{aligned}$$

If H_p is true (i.e. the PoI is the offender), there are three components to consider:

- Stain of profile Γ_1 is from the offender. Corresponding profiles are observed with probability $1 \times r_1$.

- Stain of profile Γ_2 is from the offender. The probability of corresponding profiles is zero since the PoI is assumed to be the offender and only one offender is assumed $(0 \times r_2)$.
- Neither stain is from the offender. This event has probability $(1 - r_1 - r_2)$ and if it is true there is a probability γ_1 of a correspondence between the PoI's profile (Γ_1) and the crime stain of the same profile. The probability of the combination of these events is $(1 - r_1 - r_2)\gamma_1$.

These three components are mutually exclusive and the probability in the numerator of the likelihood ratio is the sum of these three probabilities: $r_1 + (1 - r_1 - r_2)\gamma_1$.

If H_d is true (i.e. the PoI is not the offender), the probability of a correspondence is γ_1, as before.

The likelihood ratio is then

$$V = \frac{r_1 + (1 - r_1 - r_2)\gamma_1}{\gamma_1}.$$

Certain special cases can be distinguished. As r_1 and r_2 tend to zero, which implies that neither stain is relevant, then the likelihood ratio tends to 1. A likelihood ratio of 1 provides no support for either proposition, a result in this case that is entirely in agreement with the information that neither stain is relevant.

For $r_1 = r_2 = 1/2$, the value of the evidence becomes $V = 1/2\gamma_1$. For $r_1 = 1$, V equals $1/\gamma_1$. As $r_2 \to 1$, then $r_1 \to 0$ and $V \to 0$. All of these are perfectly reasonable results.

6.3.3 Paternity and The Combination of Likelihood Ratios

A special situation in which offence level propositions apply, in particular in civil proceedings, is that of paternity cases. At trial the plaintiff's proposition will generally be the allegation of a woman that the PoI is the father of her child. The alternative proposition, that of the PoI, is that he is not the father of the child. Thus, in paternity cases, the likelihood ratio is used to compare two probabilities, as in the criminal context. More formally, the propositions of interest are

H_p: the alleged father is the genetical father,

H_d: an unknown person is the genetical father.

The probability of seeing the child's non-maternal alleles given they were passed by the alleged father is compared with the probability of seeing those alleles if they came from an unknown man. Thus the value V of the evidence is

$$V = \frac{\text{Probability of the alleles given they were passed by the alleged father}}{\text{Probability of the alleles given they were passed by a unknown man}}.$$

This ratio has also been called the *paternity index* (PI) by Salmon and Salmon (1980). The example discussed here is an example of the use of the likelihood ratio to include more than one item of

Table 6.9 Allelic configuration of three individuals, a child, a mother, and an alleged father, regarding two DNA markers in a hypothetical case of disputed paternity.

Evidence	Locus	Genotypes		
		Child	Mother	Alleged father
E_1	Penta D	13–13	9–13	11–13
E_2	VWA	18–19	16–19	18–18

evidence through consideration of more than one DNA marker.

For illustration, consider two items of evidence E_1 and E_2, referring to the two genetic markers (*loci*) Penta D and VWA.[4] DNA profiling results are available for the child, the mother and the alleged father as shown in Table 6.9.

The likelihood ratio, or paternity index PI, for $E_i, i = 1, 2$, is

$$\frac{\Pr(E_i \mid H_p)}{\Pr(E_i \mid H_d)}.$$

Let G_{Ci}, G_{Mi}, and G_{AFi} denote the genotypes of the child C, the mother M, and the alleged father AF, respectively, for evidence E_i. Let A_{Mi} and A_{Pi} denote the maternal and paternal alleles for evidence E_i. Let $\gamma_{i,j}$ be the population proportion of allele j for evidence E_i.

[4]Scientists generally use commercial kits offering, for example, results for 16 markers simultaneously.

For E_1, the numerator of the likelihood ratio, $\Pr(G_{C1} \mid G_{M1}, G_{AF1}, H_p)$, is $1/4$. This is because parents with genotypes 9–13 and 11–13, respectively, will have a child with genotype 13–13 with probability $1/4$. The denominator is given by:

$$\begin{aligned}
&\Pr(G_{C1} \mid G_{M1}, G_{AF1}, H_d) \\
&\quad = \Pr(A_{M1} \mid G_{M1}) \times \Pr(A_{P1} \mid H_d) \\
&\quad = \Pr(A_{M1} = 13 \mid G_{M1} = 9\text{–}13) \\
&\qquad \times \Pr(A_{P1} = 13 \mid H_d) \\
&\quad = (1/2) \times \gamma_{1,13}.
\end{aligned}$$

The likelihood ratio for Penta D is then $1/(2\gamma_{1,13})$.

For E_2, the numerator of the likelihood ratio, $\Pr(G_{C2} \mid G_{M2}, G_{AF2}, H_p)$, equals $1/2$, because a couple with genotypes 16–19 and 18–18, respectively, will have a child with genotype 18–19 with probability $1/2$. The denominator is

$$\begin{aligned}
&\Pr(G_{C2} \mid G_{M2}, G_{AF2}, H_d) \\
&\quad = \Pr(A_{M2} \mid G_{M2}) \times \Pr(A_{P2} \mid H_d) \\
&\quad = \Pr(A_{M2} = 19 \mid G_{M2} = 16 - 19) \\
&\qquad \times \Pr(A_{P2} = 18 \mid H_d) \\
&\quad = (1/2) \times \gamma_{2,18}.
\end{aligned}$$

The likelihood ratio for VWA is then $1/\gamma_{2,18}$.

Under an assumption of independence (see Section 6.1.7 for a comment on this aspect), the likelihood ratio for the combination (E_1, E_2) of

evidence is

$$\frac{\Pr(E_1, E_2 \mid H_p)}{\Pr(E_1, E_2 \mid H_d)} = \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(E_2 \mid H_p)}{\Pr(E_2 \mid H_d)} = \frac{1}{2\gamma_{1,13}} \times \frac{1}{\gamma_{2,18}}.$$

Assume that $\gamma_{1,13} = 0.206$ and $\gamma_{2,18} = 0.227$ in a given relevant population. Then $1/(2\gamma_{1,13}\gamma_{2,18}) = 10.7 \simeq 11$. Thus, the evidence of the two marker systems is approximately 11 times more probable given the proposition that the alleged father is the true father than the proposition that an unknown male is the father. From Table 2.8 this represents moderate support for the proposition that the alleged father is the true father. It is this ratio that is known as the paternity index.

If the assumption of independence is not accepted and subpopulation effects have to be considered, the likelihood ratio takes a different form. For illustration, consider the second marker and the probability $\Pr(G_{C2} \mid G_{M2}, G_{AF2}, H_d)$. Here, $\Pr(A_{P2} \mid G_{M2}, G_{AF2}, H_d)$ cannot be reduced to $\Pr(A_{P2} \mid H_d)$, because the observed alleles should be taken into account. Assume that the mother, the alleged father and the true father all belong to the same population, then the mother's and the father's alleles are conditioning the probability of the true father's alleles: $\Pr(A_{P2} = 18 \mid G_{M2} = 16-19, G_{AF2} = 18-18, H_d)$. Paternal allele 18 is therefore conditioned by the previous observation of alleles 16, 19, and 18. Using the

formula of Balding and Nichols (1994) (6.13) for the probability that amongst n alleles drawn randomly from the subpopulation, the first n_i are of a given type of interest and the following $n - n_i$ are of other types, the paternity index, for the two markers, becomes

$$\text{PI} = \frac{1 + 3\theta}{2[2\theta + (1 - \theta)\gamma_{1,13}]} \times \frac{1 + 3\theta}{2\theta + (1 - \theta)\gamma_{2,18}},$$

where θ represents the co-ancestry coefficient.

Consider another case and, again, for the sake of illustration, marker 2 only. Assume that only the father and the alleged father belong to the same subpopulation, but not the mother. The PI will change again due to the fact that only the alleles of the alleged father are used as a conditioning, i.e. $\Pr(A_{P2} = 18 \mid G_{AF2} = 18 - 18, H_d)$. The PI thus becomes

$$\text{PI} = \frac{1 + \theta}{2\theta + (1 - \theta)\gamma_{2,18}}.$$

Note that generally forensic scientists use formulae for calculating the probability of DNA profiles for two related individuals under an assumption of independence of genes. Balding and Nichols (1995) consider paternity indices for the case where the mother, alleged father and alternative father all belong to the same sub-population and values of the allele proportions are available only for the total (general) population. Paternity formulae considering parentage or other

alleged relationships, when only two individuals are analysed (i.e. alleged parent and child), are proposed by Ayres (2000) and Lee et al. (2000). The formulae incorporate the co-ancestry coefficient, F_{ST}. The effect of incorporating F_{ST} into the equations is, in most situations, to decrease the paternity index for parentage. In fact, uncertainty arises from the fact that corresponding profiles could be due to allele sharing between the alleged father and the set of alternatives as specified by H_d.

Formulae for the PI have also been developed for some of the most often encountered alleged relationships between two individuals, such as alleged full siblings and half-siblings versus unrelated individuals (Fung et al., 2003). Mutation probabilities have also been incorporated into likelihood ratios. For a discussion on mutation rates and their assignment (see Dawid et al. (2001), Vicard and Dawid (2003), Vicard et al. (2008), Slooten and Ricciardi (2013)).

Moreover, it has been shown that it is important to allow for the fact that a close relative of the alleged father may be the true father, in addition to the usual alternative of an unrelated man. Formulae to allow for this are given in Lee et al. (1999).

Likelihood ratio equations have also been developed for cases in which additional family members are analysed or the alleged father is deceased (Evett and Weir, 1998). Mortera et al. (2016) and Green and Mortera (2017) considered situations involving DNA mixtures. A general formula

for a likelihood ratio that is appropriate for the evaluation of many potential relationships, based on two DNA profiles, is presented in Brenner and Weir (2003). Fung (2003) proposed a computerised approach for the calculation of likelihood ratios. Bayesian networks have been developed and proposed for situations involving additional family typings, absence of members of a family or complicated pedigrees (e.g. Corradi et al., 2003; Cavallini and Corradi, 2006; Hepler and Weir, 2008; Taroni et al., 2014a). Mathematical solutions are given in Fung and Hu (2008) and Buckleton et al. (2016a,b,d).

6.3.4 Probability of Paternity

In the context of paternity, it is appropriate to make a digression from consideration solely of the likelihood ratio and to consider the (posterior) probability that the alleged father is the true father; i.e. the probability that H_p is true, known as the *probability of paternity*.

Consider the two items of evidence, E_1 and E_2 of Table 6.9. First, the odds in favour of H_p, given E_1 may be written, using the odds form of Bayes' theorem (2.7), as

$$\frac{\Pr(H_p \mid E_1)}{\Pr(H_d \mid E_1)} = \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

and

$$\Pr(H_d \mid E_1) = 1 - \Pr(H_p \mid E_1)$$

so

$$\Pr(H_p \mid E_1) = \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)} \times \{1 - \Pr(H_p \mid E_1)\}$$

and

$$\Pr(H_p \mid E_1) \left\{ 1 + \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)} \right\} = \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)},$$

so that

$$\Pr(H_p \mid E_1) = \left\{ 1 + \frac{\Pr(E_1 \mid H_d)}{\Pr(E_1 \mid H_p)} \times \frac{\Pr(H_d)}{\Pr(H_p)} \right\}^{-1}, \tag{6.40}$$

a result analogous to (3.6).

Suppose that the alleged father and only one other man, of unknown genetical type, could be the true father and that, initially, each of these two possibilities is considered equally probable (Essen-Möller, 1938). Then $\Pr(H_p) = \Pr(H_d) = 0.5$. More generally, note however that the assignment of a probability of 0.5 is not a necessary consequence of the assumption of there being only two possible fathers. It is perfectly feasible to entertain a prior probability of 0.5 for the alleged father being the true father, given the background information of the case at hand, even when there is more than one alternative father. Using again the example introduced in Section 6.3.3, the posterior probability of the alleged father being the true father is

obtained as follows:

$$\Pr(H_p \mid E_1) = 1/(1 + 2\gamma_{1,13}) = 1/(1 + 0.412) = 0.708.$$

Now, second, include E_2. The posterior odds $\Pr(H_p \mid E_1)/\Pr(H_d \mid E_1)$ in favour of H_p, given E_1, replace the prior odds $\Pr(H_p)/\Pr(H_d)$ (5.22) and the posterior probability for H_p, given E_1 and E_2, is given by

$$\begin{aligned}
&\Pr(H_p \mid E_1, E_2) \\
&= \left\{ 1 + \frac{\Pr(H_d \mid E_1)}{\Pr(H_p \mid E_1)} \times \frac{\Pr(E_2 \mid H_d)}{\Pr(E_2 \mid H_p)} \right\}^{-1} \\
&= \left(1 + \frac{0.292}{0.708} \times \frac{0.227/2}{1/2} \right)^{-1} \\
&= 0.914, \qquad (6.41)
\end{aligned}$$

where the assumption of the independence of E_1 and E_2 has been made just to simplify the notation. The probability that the alleged father is the true father was, initially, 0.5. After presentation of the Penta D evidence, E_1, it became 0.708. After the presentation of the VWA evidence, E_2, it became 0.914. Note that this posterior probability is just the ratio of PI to (1 + PI), i.e. in the case here, $10.7/11.7 = 0.914$.

Note that it is important to properly define the space of propositions in order to avoid breaches of the laws of probability (Berry, 1991; Allen et al., 1995). For example, if there are two fathers, both

of type $(11-13, 18-18)$, it is not appropriate then to calculate a posterior probability by considering the alleged father 1 versus an unrelated male, and then another posterior probability by considering alleged father 2 versus an unrelated male. This would lead, for each of the two alleged fathers, to a posterior probability of 0.914 of being the true father. The probability of one or other being the true father would then be the sum of these two probabilities, i.e. 1.828. However, this is greater than 1 and breaches the first law of probability (1.6). Clearly, thus, if there is more than one alleged father, they must be considered within the same space of propositions, such as alleged father 1 is the true father versus alleged father 2 is the genetical father including, if necessary, also the proposition that an unrelated male is the true father. Further discussion of such intricacies are given in Ellman and Kaye (1979), Kaye (1989), Allen et al. (1995), and Taroni and Aitken (1998a). This situation involves multiple propositions and is approached as presented in Section 6.1.6.2.

Concerning prior probability assignment, note that, for a long time, some Courts have been aware of the unrealistic nature of default assumptions, such as equal prior probabilities. The following is an example from *Re the Paternity of M.J.B. : T.A.T.* (1988):

> *Leaving the choice of the prior odds to the legal decision maker is preferable to presenting or using an unarticulated prior probability.*

Table 6.10 Posterior probabilities of paternity for various prior probabilities for evidence for alleged father, $E_1 = 11–13$ and $E_2 = 18–18$.

$\Pr(H_p)$	0.5	0.25	0.1	0.01
$\Pr(H_p \mid E_1)$	0.708	0.447	0.195	0.024
$\Pr(H_p \mid E_1, E_2)$	0.914	0.781	0.516	0.097

The effect on the posterior probability of altering the prior probability can be determined from (6.40) and (6.41). Some sample results are given in Table 6.10.

This idea has been repeatedly expressed, for example, in the case *State of New Jersey vs. J.M. Spann* (1993):

> *The expert's testimony should be required to include an explanation to the jury of what the probability of paternity would be for a varying range of such prior probabilities running from 0.1 to 0.9.*

The International Society of Forensic Genetics, in its recommendations on biostatistics in paternity testing, emphasises (Gjertson et al., 2007):

> *In addition to the combined PI, test reports shall also contain the individual PI's for each genetic system reported and the racial/ethnic backgrounds used by the laboratory for calculations. If the probability of paternity (W) is reported, then the prior probability assumption used to calculate W shall be stated. Test reports shall include*

> *statements of assumptions, validation and computational techniques whenever alternative biostatistical methods to PI are used. (p. 228)*

A general discussion about prior probabilities, in particular equal prior probabilities, can be found in Biedermann et al. (2007).

The probability $\Pr(H_p \mid E_1, E_2)$ for independent E_1 and E_2 may be written as

$$\left\{ 1 + \frac{\Pr(E_1 \mid H_d)}{\Pr(E_1 \mid H_p)} \times \frac{\Pr(E_2 \mid H_d)}{\Pr(E_2 \mid H_p)} \times \frac{\Pr(H_d)}{\Pr(H_p)} \right\}^{-1}.$$

and if $\Pr(H_d)$ and $\Pr(H_p)$ are taken equal to 0.5 then

$$\Pr(H_p \mid E_1, E_2) = \left\{ 1 + \frac{\Pr(E_1 \mid H_d)}{\Pr(E_1 \mid H_p)} \times \frac{\Pr(E_2 \mid H_d)}{\Pr(E_2 \mid H_p)} \right\}^{-1}.$$

In general, for n independent DNA markers, representing evidence $E_1, E_2, \ldots, E_n$, with $\Pr(H_p) = \Pr(H_d)$,

$$\Pr(H_p \mid E_1, \ldots, E_n) = \left\{ 1 + \prod_{i=1}^{n} \frac{\Pr(E_i \mid H_d)}{\Pr(E_i \mid H_p)} \right\}^{-1}$$

where $\prod_{i=1}^{n} \Pr(E_i \mid H_d) / \Pr(E_i \mid H_p)$ is the product of the reciprocals of the n likelihood ratios $\Pr(E_i \mid H_p) / \Pr(E_i \mid H_d)$. This expression is called the *plausibility of paternity* (Berry and Geisser,

1986). Notice that it depends on the assumption $\Pr(H_p) = \Pr(H_d) = 0.5$, which is a default assumption and unrealistic in many cases. The assumption that $\Pr(H_p)$ equals $\Pr(H_d)$ may easily be dispensed with to give the following result

$$\Pr(H_p \mid E_1, \ldots, E_n) = \left\{ 1 + \frac{\Pr(H_d)}{\Pr(H_p)} \prod_{i=1}^{n} \frac{\Pr(E_i \mid H_d)}{\Pr(E_i \mid H_p)} \right\}^{-1}.$$

The plausibility of paternity is also sometimes referred to as the *likelihood of paternity* (Hummel, 1971, 1983) to provide a verbal scale, given here in Table 6.11. Notice that this verbal scale is one for probabilities (on the propositions). Hummel (1983) provided other values for the ranges of plausibility of paternity and corresponding verbal equivalent. The Hummel (1971) verbal scale has been adopted by Swiss jurisprudence. Note also that the verbal scale provided by Table 2.8 is for likelihood ratios.

6.3.5 Absence of Evidence for Offence Level Propositions

This is an example where a Bayesian network structure can be elicited from an existing likelihood ratio formula (Taroni et al., 2004). The problem of interest is missing evidence, for which

Table 6.11 Likelihood of paternity.

Plausibility of paternity	Likelihood of paternity
0.9980–0.9990	Practically proved
0.9910–0.9979	Extremely likely
0.9500–0.9909	Very likely
0.9000–0.9499	Likely
0.8000–0.8999	Undecided
0.8000	Not useful

Source: Based on Hummel (1971).

Lindley and Eggleston (1983) have provided a general Bayesian formula. According to Schum (1994) (as discussed in Section 6.2.6), evidence is called *missing* if it is expected, but is neither found nor produced on request. The example presented in Lindley and Eggleston (1983) relates to a collision between two motor cars, described as follows:

> *The plaintiff sues the defendant, claiming that it was his car that collided with the plaintiff's. The evidence (...) is weak, and the defendant relies on the fact that, his car being red, the plaintiff has produced no evidence that any paint, red or otherwise, was found on the plaintiff's car after the collision. (p. 87)*

A likelihood ratio to assist the court in the examination of the effect that the evidence is missing (variable M) has on the truth or otherwise of the main proposition of interest (variable H) is

(Lindley and Eggleston, 1983):

$$\frac{\Pr(M \mid H_p)}{\Pr(M \mid H_d)} = \frac{\begin{array}{r}\Pr(M \mid E_1)\Pr(E_1 \mid H_p) \\ +\Pr(M \mid E_2)\Pr(E_2 \mid H_p) \\ +\Pr(M \mid E_3)\Pr(E_3 \mid H_p)\end{array}}{\begin{array}{r}\Pr(M \mid E_1)\Pr(E_1 \mid H_d) \\ +\Pr(M \mid E_2)\Pr(E_2 \mid H_d) \\ +\Pr(M \mid E_3)\Pr(E_3 \mid H_d)\end{array}}. \tag{6.42}$$

The variable E designates the form of the evidence that is missing, defined as follows: there was red paint on the plaintiff's car (E_1), there was paint on the plaintiff's car, but it was not red (E_2), there was no paint on the plaintiff's car (E_3).

The construction of a Bayesian network based upon an existing formula has the advantage that the number and definition of the nodes is already given. From the above example on missing evidence, three variables are of interest.

(1) The variable H represents the event that the defendant committed the offence for which he has been charged. This event may either be true of false, represented by the two states H_p and H_d.

(2) The variable M represents the event that evidence is missing. This variable can take the value true or false, denoted M and $\bar{M}$, respectively.

(3) The variable E designates the form of the evidence that is missing, with three states E_1, E_2, and E_3 as defined above.

In order to find a graphical representation that appropriately reflects the conditional dependencies as specified by the likelihood ratio displayed above (6.42), it is helpful to follow a two-stage approach. Lindley and Eggleston (1983) assert that their formula (6.42) contains all the relevant considerations for the paint case, notably conditional probabilities for:

- the various forms of the evidence given that the prosecution hypothesis H_p is true and given the defence hypothesis H_d is true,
- the evidence being missing were it E_1, E_2, or E_3, respectively.

Consider the first of the two points mentioned earlier. Accepting that the probability of the evidence is conditioned on the truth state of the variable H means, graphically, that H is chosen as a parental variable for E, as shown in Figure 6.2a. One can proceed similarly for the second of the above two points. If the event that evidence is missing (M) is conditioned on the form the missing evidence can take (E), then E can be chosen as a parental variable for M (see Figure 6.2b). Next, since the variable E shown in Figure 6.2a is the same as in Figure 6.2b, the two network fragments combine to give the Bayesian network structure shown in Figure 6.2c.

When searching for a sound structure for a Bayesian network, based on the three variables M, E, and H, it is legitimate to ask whether there could be an arrow pointing from H to M. Consideration

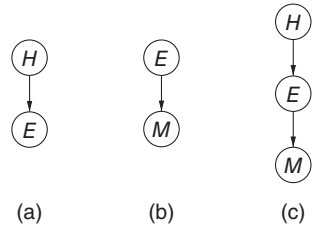

Figure 6.2 Bayesian network fragments for the problem of missing evidence, representing the relation between (a) the variables E and H, and (b) the variables M and E respectively; (c) Bayesian network for missing evidence.

of the proposed network structure as shown in Figure 6.2c indicates that there should be none. As with any other graphical element employed in Bayesian network structures, the absence of an arrow must be justified as well. In the current example, the absence of a directed edge between H and M can be justified by the argument given by Lindley and Eggleston (1983), who assumed that

> *(. . .) were the form of the missing evidence known, then the view of the defendant's guilt would not be altered by knowing that this evidence had, or had not, been produced in court. (p. 90)*

Stated otherwise '[...] the actual evidence suppresses any importance being attached to its omission' (p. 91). In formal notation, this corresponds to the following: $\Pr(H_p \mid E, M) = \Pr(H_p \mid E)$, where E may take any of its three possible states E_1, E_2, or E_3, and M either be M or $\bar{M}$. The proposed Bayesian network correctly encodes this property through its serial connection, where H and M are independent conditional on knowing

E. Practically, this means that the transmission of evidence between the nodes H and M is blocked whenever E is instantiated. This is also sometimes expressed as the node E 'screening off' M from H, which is an example of d-separation.

The main aim of the current example is to discuss, using graphical models, the logical structure of the proposed likelihood ratio for missing evidence. Thus, the numerical specification of the Bayesian network for the current example is not covered here. However, the implementation of the Bayesian network in a suitable Bayesian network software would provide further means to validate the proposed network structure. In particular, the effect of different probabilities for the suppression of favourable and unfavourable evidence for the defendant on the odds of the offence for which he has been charged can be examined and compared with the indications given in Lindley and Eggleston (1983).

6.3.6 A Note on Relevance and Offence Level Propositions

Relevance is a factor to be taken into account when assessing findings given offence level propositions, but not when considering activity level propositions. Let γ be the population proportion for an analytical trait or feature. Under offence level propositions, the value of the evidence cannot be greater than $1/\gamma$ (if $k = 1$, $r = 1$ and $p = 0$). It may be reduced below this value because relevance, expressed as a probability, may be less

than 1 and the probability the crime stain comes from the PoI given that they did not commit the crime and the crime stain did not come from the offender(s) is not equal to 0. Under activity level propositions, the value of the evidence can be greater than $1/\gamma$. Here, factors such as transfer and background presence have to be considered.

However, under offence level propositions and assuming that the PoI has to be present to commit the crime, then the activity level does not need to be considered. Consider two propositions, H_{p1} that the PoI committed the crime (offence level) and H_{p2} that the PoI was present at the crime scene (activity level). Then

$$\Pr(E \mid H_{p1}, H_{p2}) = \Pr(E \mid H_{p1}).$$

6.4 SUMMARY

The main results of the previous sections may usefully be summarised.

6.4.1 Stain Known to Have Been Left by Offenders: Source-Level Propositions

6.4.1.1 One Stain Known to Have Come from One Offender

The profile of the crime stain and the profile of the PoI is Γ. The proportion of the relevant population that has type Γ is γ. The propositions to be compared are

H_p: the stain at the crime scene came from the PoI;

H_d: the stain at the crime scene came from an unknown person.

Then

$$V = \frac{1}{\gamma}.$$

6.4.1.2 Two Stains, One from Each of Two Offenders

There are two crime stains, of profiles Γ_1 and Γ_2, with associated population proportions γ_1 and γ_2. There is one PoI with profile Γ_1. The propositions to be compared are

H_p: the crime stains came from the PoI and one other person;

H_d: the crime stains came from two unknown people.

Then

$$V = \frac{1}{2\gamma_1}.$$

6.4.1.3 Multiple (n) Stains, One from Each of n Offenders

There is one PoI with profile Γ_1 and associated with population proportion γ_1. The propositions to be compared are

H_p: the crime stains came from the PoI and $(n-1)$ other people;

H_d: the crime stains came from n unknown people.

Then

$$V = \frac{1}{n\gamma_1}.$$

6.4.1.4 Multiple (n) Stains, k Different Profiles, k Different Offenders

There are s_i stains of type i ($i = 1, \ldots, k$; $\sum_{i=1}^{k} s_i = n$). There is one PoI with profile Γ_1 and associated population proportion γ_1. The propositions to be compared are

H_p: the crime stains came from the PoI and $(k - 1)$ other people;

H_d: the crime stains came from k unknown people.

Then

$$V = \frac{1}{\binom{n}{s_1}\gamma_1^{s_1}}.$$

6.4.1.5 Multiple (n) Stains, k Different Profiles, m Offenders

This situation may arise where there are limited analytical results. There are m_i offenders with profile i ($i = 1, \ldots, k$; $\sum_{i=1}^{k} m_i = m$). There are s_{ij} stains belonging to the jth offender for the ith profile. There is one PoI with blood group Γ_1 and associated population proportion γ_1. Assume without loss of generality this is the first offender with the first profile. The propositions to be compared are

H_p: the crime stains from the PoI and $(m-1)$ other people;

H_d: the crime stains came from m other people.

Then

$$V = \frac{1}{\binom{n}{s_{11}} \gamma_1^{s_{11}}}.$$

6.4.2 Material Known to Have Been (or Not to Have Been) Left by Offenders: Activity-Level Propositions

6.4.2.1 Material Left by The Offender

In cases involving an activity, the scientist may encounter a situation where the transfer material has been left by the offender. The propositions to be compared are

H_p: the PoI assaulted the victim;

H_d: an unknown person assaulted the victim.

Then

$$V = \frac{b_0 t_n + b_{g,\mathbf{m}} \gamma t_0}{b_0 \gamma t'_n + b_{g,\mathbf{m}} \gamma t'_0}.$$

6.4.2.2 Material Not Left by The Offender

When the recovered material has not been left by the offender, the likelihood ratio is reduced to

$$V = \frac{b_0 t_n + b_{g,\mathbf{m}} \gamma t_0}{b_{g,\mathbf{m}} \gamma},$$

6.4.2.3 Innocent Transfer of Material

In cases where an innocent transfer can be assumed, the likelihood ratio is

$$V = \frac{b_0 tc + b_{1,m}\gamma[(1-t)c + (1-c)]}{b_0 sd + b_{1,m}\gamma[(1-s)d + (1-d)]},$$

where c is the probability the victim has been in contact with the PoI given proposition H_p, d is the probability the PoI could have been in contact with the victim for reasons other than the assault, and s represents the transfer probability given the alternative proposition H_d; see Section 5.3.2.3.

6.4.2.4 Material Left by The Offender (Uncertainty About The True Source)

In situations involving material, which is not unique to one source (such as an individual's DNA) but can be shared as fibres from pullovers, shoeprints, and so on, there is uncertainty about the assumed known source. Therefore, the likelihood ratio can be expressed as

$$V = \frac{b_0 t_n \delta + b_{g,\mathbf{m}}\gamma[t_0\delta + t_0''](1-\delta)]}{b_0\gamma t_n' + b_{g,\mathbf{m}}\gamma t_0'}, \qquad (6.43)$$

where δ refers to $[w + \gamma'(1-w)]$, w specifies the probability that the PoI wore the known source, and t_0'' represents the probability of no transfer from the true source; see (5.9).

6.4.3 Stain May Not Have Been Left by Offenders: Offence-Level Propositions

6.4.3.1 One Stain, k Offenders

Relevance is defined as the probability that a crime stain came from one of the k offenders. This probability is denoted r. The propositions to be compared are

H_p: the PoI is one of the k offenders;

H_d: the PoI is not one of the k offenders.

The stain is of profile Γ. It may have been left by an offender. There are k offenders. The PoI is of profile Γ. The applicable population proportion for this profile is γ, referring to the population from which the criminals may be thought to have come. The population proportion is γ' amongst the people who may have left the stain that may not be the same population as that from which the criminals may be thought to come. For example, there may be eyewitness evidence that the criminals are from one ethnic group whereas the people normally associated with the crime scene may be from another. The probability that the stain would have been left by the PoI even though they were innocent of the offence is p.

$$V = \frac{[r\{1 + (k-1)\gamma\}/k] + \{\gamma'(1-r)\}}{\gamma r + \{p + (1-p)\gamma'\}(1-r)}$$

$$= \frac{r\{1 + (k-1)\gamma\} + k\gamma'(1-r)}{k[\gamma r + \{p + (1-p)\gamma'\}(1-r)]}.$$

In some situations, the likelihood ratio reduces to simpler forms: For example, if $\gamma = \gamma'$ and $p = 0$ then

$$V = \frac{r + (k-r)\gamma}{k\gamma}.$$

If, also, $r = 1$,

$$V = \frac{1 + (k-1)\gamma}{k\gamma}.$$

Compare this with the case in which there are n stains (rather than 1) and the number (k) of offenders equals the number of stains (n), with one stain coming from each of n offenders. Then $V = 1/n\gamma = 1/k\gamma$.

If there is one stain and k offenders

$$V = \frac{1 + (k-1)\gamma}{k\gamma}$$
$$= 1 + \frac{1}{k\gamma} - \frac{1}{k},$$

and an increase of the value of the likelihood ratio is observed compared with the situation $n = k$ where $n \neq 1$. However, if γ is small, V becomes, independently of the scenario, approximately equal to $1/k\gamma$.

If $\gamma = \gamma'$ and $p \neq 0$ then

$$V = \frac{[r\{1 + (k-1)\gamma\}/k] + \{\gamma(1-r)\}}{[\gamma r + \{p + (1-p)\gamma\}(1-r)]}$$

$$= \frac{r + (k - r)\gamma}{k[p(1 - r) + \gamma\{r + (1 - p)(1 - r)\}]}$$
$$= \frac{r + (k - r)\gamma}{k[p(1 - r) + \gamma(1 - p + pr)]}.$$

6.4.3.2 Two Stains, One of Which is Relevant, One Offender

The offender left one of the bloodstains but it is not known which one. The propositions to be compared are

H_p: the PoI is the offender;

H_d: an unknown person is the offender.

A PoI is of profile Γ_1, with associated population proportion γ_1. Let r be the probability that the crime stain that comes from the offender and corresponds to the group of the PoI is from the PoI. Then $(1 - r)$ is the probability that the crime stain which does not correspond to the profile of the PoI is from the offender. Then

$$V = \frac{r}{\gamma_1}.$$

6.4.3.3 Two Stains, Neither of Which May Be Relevant, One Offender

The propositions to be compared are

H_p: the PoI is the offender;

H_d: an unknown person is the offender.

A PoI is of profile Γ_1, with associated population proportion γ_1. Let r_1 and r_2 be the probabilities that

the stain of profile Γ_1 and the stain of profile Γ_2 is from the offender, respectively. Then $(1 - r_1 - r_2)$ is the probability that neither stain is from the offender and

$$V = \frac{r_1 + (1 - r_1 - r_2)\gamma_1}{\gamma_1}.$$

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

7

Data Analysis

7.1 INTRODUCTION

The evaluation of scientific evidence is often thought of as the assessment of a comparison between evidential material whose source is unknown (i.e. recovered material) and evidential material whose source is known (i.e. control material). The quantitative part of the evidence is represented by the measurements of the characteristic of interest. Let x denote a measurement on the control material and let y denote a measurement on the recovered material. These measurements can be either discrete, such as DNA profiles or the types of resins detected on black toner found on questioned documents, or continuous, such as measurements on the refractive indices of fragments of glass that are available for comparison. For example, if a window is broken during the commission of a crime, the measurements on the refractive indices of m fragments of glass found at the crime scene will be denoted $\mathbf{x} = (x_1, \ldots, x_m)^T$. The refractive indices of n fragments of glass found

on a PoI will be denoted $\mathbf{y} = (y_1, \ldots, y_n)^T$. By convention, vectors are denoted in bold font and the elements of a vector are written in a column. The corresponding row vector is denoted with a superscript^T to denote transpose (of a column to a row). See Section B.1.2 for further details. The quantitative part of the evidence concerning the glass fragments in this case can be denoted by

$$E = (\mathbf{x},\mathbf{y}).$$

In the notation of Section 1.7.1, M_c is the broken window at the crime scene, M_r is the set of glass fragments from the PoI, E_c is $\mathbf{x}$, E_r is $\mathbf{y}$, M is (M_c, M_r), and $E = (E_c, E_r) = (\mathbf{x}, \mathbf{y})$.

Let H_p and H_d denote the prosecution and the defence propositions according to which the control and recovered items originate from the same or from different sources, respectively, and I denote the background information. Then the value V of the evidence is, formally,

$$V = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}, \tag{7.1}$$

as before (see (2.15) in Section 2.4.1 and (5.1) in Section 5.3.1).

As well as the problem of comparison, the likelihood ratio may also be used when the scientist wishes to assign an observation to one of several populations on the basis of available measurements of some attributes, a procedure known

as discrimination. This situation is discussed in Section 7.7. For example, assume there are two populations, P_1 and P_2. The propositions for the likelihood ratio are H_1 and H_2, denoting membership of populations 1 and 2, respectively. Denote the measurements by $\mathbf{z}$. The likelihood ratio is then

$$\frac{\Pr(\mathbf{z} \mid H_1)}{\Pr(\mathbf{z} \mid H_2)},$$

where values of the likelihood ratio greater than one support membership of population 1 and less than one support membership of population 2. For more than two populations, prior probabilities for population membership are required in a formulation described in Section 7.7.

Many of the models described in the chapter are Bayesian hierarchical random effects models. A review of these models in forensic science is given in Aitken (2018).

7.2 THEORY FOR DISCRETE DATA

Several proposals and examples can be found in the literature when data are continuous. This is not so when data are discrete. Other than for DNA profiling there is a paucity of methods and proposals when data are discrete. Consider the case where discrete measurements, such as counts, are collected in correspondence of recovered and

control material that is available for comparison. The likelihood ratio (7.1) may be written as

$$V = \frac{\Pr(X = x, Y = y \mid H_p, I)}{\Pr(X = x, Y = y \mid H_d, I)}. \qquad (7.2)$$

Among problems and difficulties that may be encountered, there are the numbers of probabilities that need to be elicited. If several variables are measured (say, p) and each variable has several levels of response (say, k for each variable), there will be $(k-1)^p$ marginal probabilities to be assessed along with interactions, with appropriate adjustments if there are differing levels of responses for each variable. Independence between variables allows elementary models for counts to be applied, but this is not always feasible. In what follows, a likelihood ratio will be computed for observations in the form of independent counts from a Poisson distribution (Section 7.2.1) and for observations in the form of realizations of independent trials in which the target characteristic will take one of two or more mutually exclusive outcomes (Section 7.2.2 and 7.2.3).

Another problem that may be encountered is related to the possible autocorrelation between adjacent observations. For example, it is possible that drug traces are transferred from a contaminated banknote to an adjacent one. A model that takes into account autocorrelation has been proposed by Wilson et al. (2014, 2015) for continuous data and will be discussed briefly in Section 7.7.3.

7.2.1 Data of Independent Counts with a Poisson Distribution

Consider the following scenario where a crime is committed and a piece of recorded speech of unknown origin is available. The number of occurrences of a given characteristic (e.g. a click, a parameter that can be analysed in speech (Aitken and Gold, 2013)) of the speech in each of a succession of consecutive time periods, say, k_y, is noted. These are the recorded data, $\mathbf{y} = (y_1, \ldots, y_{k_y})$. Assume these characteristics are independent between different time periods and that observations can be treated as realizations from a Poisson distribution centred at λ_r (Section A.2.6). The probability distribution of available counts can be obtained as follows:

$$\begin{aligned}\Pr(\mathbf{Y} = \mathbf{y} \mid \lambda_r) &= \prod_{i=1}^{k_y} \Pr(Y_i = y_i \mid \lambda_r) \\ &= \prod_{i=1}^{k_y} \frac{\mathrm{e}^{-\lambda_r} \lambda_r^{y_i}}{y_i!} = \frac{\mathrm{e}^{-k_y \lambda_r} \lambda_r^{t_y}}{\prod_{i=1}^{k_y} y_i!},\end{aligned}$$

where $t_y = \sum_{i=1}^{k_y} y_i$.

The number of occurrences of the same characteristic of the speech of a suspect in each of a succession of consecutive time periods, say, k_x, is noted. These are the control data, $\mathbf{x} = (x_1, \ldots, x_{k_x})$, and can be treated again as realizations from a Poisson distribution centred at

λ_c. The probability distribution of available counts can be obtained analogously as

$$\Pr(\mathbf{X} = \mathbf{x} \mid \lambda_c) = \frac{\mathrm{e}^{-k_x \lambda_c} \lambda_c^{t_x}}{\prod_{i=1}^{k_x} x_i!},$$

where $t_x = \sum_{i=1}^{k_x} x_i$.

The following propositions might be of interest:

H_p: the recovered and the control speech originate from the same source;

H_d: the recovered and the control speech originate from different sources.

If available counts **y** and **x** originate from the same source (i.e. if proposition H_p is true), then $\lambda_r = \lambda_c$. *Vice versa*, if these counts do not originate from the same source, then λ_r may or may not equal λ_c. The mean number, say, λ, being unknown, a prior distribution $f(\lambda)$ will be introduced to model prior uncertainty about λ and application to the likelihood ratio in (7.2) will give

$$V = \frac{\int \Pr(\mathbf{X} = \mathbf{x} \mid \lambda) \Pr(\mathbf{Y} = \mathbf{y} \mid \lambda) f(\lambda) d\lambda}{\int \Pr(\mathbf{X} = \mathbf{x} \mid \lambda) f(\lambda) d\lambda \int \Pr(\mathbf{Y} = \mathbf{y} \mid \lambda) f(\lambda) d\lambda}. \tag{7.3}$$

The most common distribution for λ is the gamma distribution, denoted $Ga(\alpha, \beta)$, a continuous distribution parameterised by α and β, with probability density function given by

$$f(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda}; \quad \alpha > 0,\ \beta > 0,\ \lambda > 0,$$

(Section A.3.5). The numerator of the likelihood ratio in (7.3) can be computed as (see Aitken and Gold (2013))

$$\int \frac{\mathrm{e}^{-k_x\lambda_c}\lambda_c^{t_x}}{\prod_{i=1}^{k_x} x_i!}\frac{\mathrm{e}^{-k_y\lambda_r}\lambda_r^{t_y}}{\prod_{i=1}^{k_y} y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\mathrm{e}^{-\beta\lambda}d\lambda$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^{k_x} x_i!\prod_{i=1}^{k_y} y_i!} \times \frac{\Gamma(\alpha+t_x+t_y)}{(\beta+k_x+k_y)^{\alpha+t_x+t_y}}.$$

For the denominator, the marginal distribution of observations $\mathbf{X} = \mathbf{x}$ can be obtained as

$$\int \frac{\mathrm{e}^{-k_x\lambda}\lambda^{t_x}}{\prod_{i=1}^{k_x} x_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\mathrm{e}^{\beta\lambda}d\lambda = \frac{\beta^\alpha}{\prod_{i=1}^{k_x} x_i!}\frac{\Gamma(\alpha+t_x)}{(\beta+k_x)^{\alpha+t_x}}.$$

Analogously, the marginal distribution of observations $\mathbf{Y} = \mathbf{y}$ will result

$$\int \frac{\mathrm{e}^{-k_y\lambda}\lambda^{t_y}}{\prod_{i=1}^{k_y} y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\mathrm{e}^{\beta\lambda}d\lambda = \frac{\beta^\alpha}{\prod_{i=1}^{k_y} y_i!}\frac{\Gamma(\alpha+t_y)}{(\beta+k_y)^{\alpha+t_y}}.$$

The likelihood ratio is then

$$\frac{\Gamma(\alpha+t_x+t_y)\Gamma(\alpha)}{\Gamma(\alpha+t_x)\Gamma(\alpha+t_y)} \times \frac{(\beta+k_x)^{\alpha+t_x}(\beta+k_y)^{\alpha+t_y}}{\beta^\alpha(\beta+k_x+k_y)^{\alpha+t_x+t_y}}. \tag{7.4}$$

Imagine the number of clicks in each of $k_x = 6$ and $k_y = 6$ minutes of recorded speech are registered. The total number of observed clicks amounts to $t_x = 4$ and $t_y = 4$. Parameters α and β of the gamma distribution can be elicited through the method of moments. The mean of a gamma

distributed random variable is equal to α/β, and the variance is equal to α/β^2 (Section A.3.5). The prior mean and the prior variance can be equated to the sample mean and the sample variance from a training set (whenever available) and solved for α and β. Alternatively, values for the expectation and variance may be obtained through subjective elicitation. Suppose it is believed that the mean number of clicks in a given time period is equal to 3 and the variance is equal to 3, then $\alpha = 3$ and $\beta = 1$. The likelihood ratio in (7.4) then gives

$$\frac{\Gamma(3+4+4)\Gamma(3)}{\Gamma(3+4)\Gamma(3+4)} \times \frac{(1+6)^{3+4}(1+6)^{3+4}}{(1+6+6)^{3+4+4}} = 5.3,$$

a value that gives weak support (Table 2.8) to the proposition that the recovered and control speech originate from the same source.

In another application, the Poisson distribution represents a standard approach to model the number of *gunshot residues* (GSR) that might be found on an individual that shot firearms, or the number of *consecutive matching striations* (CMS) for the assessment of identification of firearms and toolmarks. A Bayesian approach was studied by Biedermann et al. (2009a), Taroni et al. (2014a), and Biedermann et al. (2011a) to the analysis of GSR and by Bunch (2000), Buckleton et al. (2008), and Bunch (2013) to the analysis of CMS data.

7.2.2 Data of Independent Counts with a Binomial Distribution

Many practical situations are encountered where available measurements are in the form of realizations of experiments that may assume only two mutually exclusive outcomes. These include DNA profiling (e.g. questioned and control material showing the same characteristic) and the presence or absence of a striation in a tool mark.

Consider the scenario described in Section 5.3.1.5 involving shoemarks recovered at a crime scene. A person of interest owns a pair of shoes producing prints (of type T) indistinguishable from the marks recovered at the crime scene. The evidence is the recovered material ($E_r = T$) and the control material ($E_c = T$). Assume the probability a shoeprint is of type T is θ. The prosecution proposition H_p is that the shoeprints and shoemarks originate from the same source and the defence proposition H_d is that the shoeprints and shoemarks originate from different sources. Conditioning on θ, the likelihood ratio is:

$$V = \frac{\Pr(E_r = T, E_c = T, \mid H_p, \theta)}{\Pr(E_r = T, E_c = T, \mid H_d, \theta)} = \frac{1}{\theta}. \tag{7.5}$$

As stated in Section 5.3.1.5, the uncertainty in θ needs to be modelled. A beta prior distribution

can be used for this purpose (as in Sections A.3.7 and 4.2). A police shoeprint database reports a total number of n prints of type T out of a database of size N and it can be used to update the beta prior parameters to $\alpha + n$ and $\beta + N - n$, respectively. Alternatively, if all available knowledge is given by the shoeprint database, the beta prior distribution can be elicited following the approach in Section 4.3.1. If the recovered shoemark and the control shoeprint originate from the same source (i.e. if H_p holds), the probability of the evidence can be obtained as

$$\begin{aligned}&\Pr(E_r = T, E_c = T \mid H_p, n, N-n)\\ &= \int_\Theta \theta\theta^{\alpha+n-1}(1-\theta)^{\beta+N-n-1}/B(\alpha+n, \beta+N-n)d\theta\end{aligned}$$

where $\Theta = [0, 1]$.

If the recovered shoemark and the control shoeprint originate from different sources (i.e. if H_d holds), the probability of the evidence can be obtained as

$$\begin{aligned}&\Pr(E_r = T, E_c = T \mid H_d, n, N-n)\\ &= \int_\Theta \theta^2\theta^{\alpha+n-1}(1-\theta)^{\beta+N-n-1}/B(\alpha+n, \beta+N-n)d\theta.\end{aligned}$$

The value of the evidence is then

$$\begin{aligned}V &= \frac{\int_\Theta \theta^{\alpha+n}(1-\theta)^{\beta+N-n-1}d\theta}{\int_\Theta \theta^{\alpha+n+1}(1-\theta)^{\beta+N-n-1}d\theta}\\ &= \frac{B(\alpha+n+1, \beta+N-n)}{B(\alpha+n+2, \beta+N-n)}\\ &= \frac{\alpha+\beta+N+1}{\alpha+n+1}.\end{aligned}$$

It can be observed that this is just the inverse of the posterior mean of θ (see Section A.3.7). In such a case the beta posterior distribution for θ is $Be(\alpha + n + 1, \beta + N - n)$. A report of such a value does not deprive the legal system of relevant information about the case. All available information, including prior uncertainty about the unknown value of θ, are encapsulated in the reported value of the likelihood ratio. Clearly, different prior probability distributions may be used, or different databases might be available, and in such a case a different value will be reported. See Taroni et al. (2016) for further discussion.

7.2.3 Data of Independent Counts with a Multinomial Distribution

In many practical situations available measurements are in the form of realizations of experiments that may assume two or more mutually exclusive outcomes.

Consider the case of questioned printed documents, where control documents originating from a known source (i.e. a printer) are available for comparison. In the case of black toner that may be found on printed documents, resins are commonly analysed by means of Fourier Infrared Spectroscopy (FTIR), the results of which (so-called IR data) may be classified into one of several mutually exclusive categories (Biedermann et al., 2009b, 2011b, 2016c).

The following propositions might be of interest:

H_p: the recovered (questioned) and the control documents originate from the same source;

H_d: the recovered (questioned) and the control documents do not originate from the same source.

Denote by E_r and by E_c the resin group contained in the toner present on the questioned document and in the toner used by the control source (printer), respectively, and by $A_1, \ldots, A_k$ the resin group categories. Suppose that the resin type contained in the questioned document corresponds to that contained in the document from the control source, say, $E_r = E_c = A_j$, $j = 1, \ldots, k$. Suppose moreover a database is available, with the records of the resin group contained in documents originating from several printers of different models. The available data in the case at hand (including the recovered evidence, the control evidence, and the background data) are given by $E_r = A_j$, $E_c = A_j$, and by the number of counts in the k categories from the background data, $n_1, \ldots, n_k$.

Assume that observations of distinct categories can be reasonably treated as independent. Denoted by θ_j the probability a resin is in group j ($j = 1, \ldots, k$) with $\sum_{j=1}^{k} \theta_j = 1$. Available counts can be treated as realizations from a multinomial

distribution (Section A.2.4). Clearly, if proposition H_p holds, then the type of resin contained in the document in question corresponds to that of the potential source, and the probability of the evidence is

$$\Pr(E_r = A_j, E_c = A_j \mid H_p, \theta_1, \ldots, \theta_k) = \theta_j.$$

Alternatively, if proposition H_d holds, the probability of the evidence is

$$\Pr(E_r = A_j, E_c = A_j \mid H_d, \theta_1, \ldots, \theta_k) = \theta_j^2.$$

The likelihood ratio can be obtained as in (7.3), where a Dirichlet prior probability distribution (Section A.3.8) will be considered for modelling probabilities for k outcomes

$$f(\theta_1, \ldots, \theta_k \mid \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \cdot \cdots \cdot \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1},$$

where $\alpha = \sum_{j=1}^{k} \alpha_i$. The probability in the numerator of the likelihood ratio can be computed as

$$\Pr(E_r = A_j, E_c = A_j \mid H_p) = \int \theta_j \frac{\Gamma(\alpha)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdot \ldots \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} \mathrm{d}\boldsymbol{\theta},$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. The probability in the denominator of the likelihood ratio can be computed analogously. Using the available background database, the parameters of the Dirichlet prior distribution can be replaced with $\alpha_j + n_j$, $j = 1, \ldots, k$, where n_j represents the number of counts in the database in category j. The likelihood ratio then becomes

$$V = \frac{\int \theta_1^{\alpha_1+n_1-1} \cdot \cdots \cdot \theta_1^{\alpha_j+n_j} \cdot \cdots \cdot \theta_k^{\alpha_k+n_k-1} d\boldsymbol{\theta}}{\int \theta_1^{\alpha_1+n_1-1} \cdot \cdots \cdot \theta_1^{\alpha_j+n_j+1} \cdot \cdots \cdot \theta_k^{\alpha_k+n_k-1} d\boldsymbol{\theta}}$$
$$= \frac{\alpha + \sum_{i=1}^{k} n_i + 1}{\alpha_j + n_j + 1}.$$

Consider the following example from Biedermann et al. (2011b). Suppose the resin groups contained in the bi-component toner present on the control and questioned printed documents are both of type *Epoxy A*, say. Table 7.1 summarises the results obtained following the analyses of the component type (magnetism) and the polymer resins (IR-category) of each of 23 samples of bi-component black toners.

Suppose a uniform Dirichlet prior distribution is taken, that is, $\alpha_1 = \cdots = \alpha_7 = 1$. The likelihood ratio is then

$$V = \frac{7 + 23 + 1}{1 + 3 + 1} = 6.2.$$

Table 7.1 Results obtained following the analyses of, respectively, the component type (magnetism) and the polymer resins (IR-category) of 23 samples of bi-component black toner.

Resin group	Counts
1. Styrene-co-acrylate	14
2. Epoxy A	3
3. Epoxy B	2
4. Epoxy C	1
5. Epoxy D	1
6. Polystyrene	1
7. Other	1

Source: Extracted from Biedermann et al. (2011b). Reprinted with permission from Elsevier.

This provides weak support for the proposition that the questioned and the control printed documents originate from the same source.

If the common resin group had been *Styrene-co-acrylate*, the likelihood ratio would be

$$V = \frac{7 + 23 + 1}{1 + 14 + 1} = 1.9$$

and if the common resin group had been *Epoxy-D*, the likelihood ratio would be

$$V = \frac{7 + 23 + 1}{1 + 1 + 1} = 10.3.$$

The change in value with change of rarity of the resin group is easily seen.

7.3 THEORY FOR CONTINUOUS UNIVARIATE DATA

Section 7.2, considered the evaluation of the likelihood ratio where the evidence was represented by discrete data with specific reference to pieces of recorded speech (Section 7.2.1), to recovered shoemarks at the crime scene (Section 7.2.2), and to analysis of the resin type of black toner from printed documents (Section 7.2.3). The values of the evidence in different contexts were derived. However, much evidence is of a form in which measurements may be taken and for which the data are continuous. The form of the statistic for the evaluation of the evidence under these circumstances is similar to that for discrete data. Many of the examples in this chapter concern the interpretation of glass evidence; a review of the statistical interpretation of such evidence is given in Curran et al. (2000) and Curran and Hicks (2009).

Continuous measurements are being considered and the probabilities Pr are therefore replaced by probability density functions f so that

$$V = \frac{f(\mathbf{x}, \mathbf{y} \mid H_p, I)}{f(\mathbf{x}, \mathbf{y} \mid H_d, I)}. \tag{7.6}$$

Bayes' theorem and the rules of conditional probability apply to probability density functions as well as to probabilities. The value, V, of the evidence (7.6) may be rewritten in the following way

$$\begin{aligned} V &= \frac{f(\mathbf{x},\mathbf{y} \mid H_p, I)}{f(\mathbf{x},\mathbf{y} \mid H_d, I)} \\ &= \frac{f(\mathbf{y} \mid \mathbf{x}, H_p, I)}{f(\mathbf{y} \mid \mathbf{x}, H_d, I)} \times \frac{f(\mathbf{x} \mid H_p, I)}{f(\mathbf{x} \mid H_d, I)}. \end{aligned}$$

The measurements $\mathbf{x}$ are those on the source, or control, object. Their distribution and corresponding probability density function are independent of whether H_p or H_d is true. Thus

$$f(\mathbf{x} \mid H_p, I) = f(\mathbf{x} \mid H_d, I)$$

and

$$V = \frac{f(\mathbf{y} \mid \mathbf{x}, H_p, I)}{f(\mathbf{y} \mid \mathbf{x}, H_d, I)}.$$

If H_d is true, it is assumed that the measurements ($\mathbf{y}$) on the recovered item and the measurements ($\mathbf{x}$) on the control item are independent. Thus

$$f(\mathbf{y} \mid \mathbf{x}, H_d, I) = f(\mathbf{y} \mid H_d, I),$$

and

$$V = \frac{f(\mathbf{y} \mid \mathbf{x}, H_p, I)}{f(\mathbf{y} \mid H_d, I)}. \tag{7.7}$$

There are scenarios where the assumption of independence between measurements of control and

recovered materials when H_d is true, is unreliable, and must be relaxed. An example was introduced in Section 6.1.5 in the discussion around DNA profiling. As a second example, consider the case of questioned signatures. Under the hypothesis that a signature is not authentic (H_d), it has to be taken into account that a forger might attempt to reproduce the features of a target (original) signature, and therefore the control and recovered measurements cannot be taken as independent.

The numerator in (7.7) is a *predictive distribution* (Section 4.6). The denominator is the so-called *marginal distribution* of the measurements on the recovered sample in the relevant population, the definition of which is assisted by I. This formulation of the expression for V shows that for the numerator the distribution of the measurements on the recovered sample, conditional on the source measurements as well as I, is considered. For the denominator, the distribution of the recovered measurements is considered over the distribution of the whole of the relevant population.

The two propositions to be compared are

H_p: the recovered sample is from the same source as the control sample;

H_d: the recovered sample is from a different source than the control sample.

First, consider H_p. The control and recovered samples are from evidence from the same source. The measurements on this source have a true

unknown value θ, say. For example, if the measurements are of the refractive index of glass, then θ denotes the mean refractive index of the window from which the fragments have been taken. For clarity, the conditioning elements H_p and I, the background information, will be omitted in the following argument. The predictive distribution $f(\mathbf{y} \mid \mathbf{x})$ may be expressed as follows

$$\begin{aligned} f(\mathbf{y} \mid \mathbf{x}) &= \int f(\mathbf{y} \mid \theta) f(\theta \mid \mathbf{x}) d\theta \\ &= \frac{\int f(\mathbf{y} \mid \theta) f(\mathbf{x} \mid \theta) f(\theta) d\theta}{f(\mathbf{x})} \\ &= \frac{\int f(\mathbf{y} \mid \theta) f(\mathbf{x} \mid \theta) f(\theta) d\theta}{\int f(\mathbf{x} \mid \theta) f(\theta) d\theta}, \end{aligned}$$

the ratio of the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ to the marginal distribution of $\mathbf{x}$. Both distributions are independent of θ, which is integrated out.

For H_d, the situation where the control and recovered samples are from different sources, it is the measurements $\mathbf{y}$ on the recovered sample that are the ones of interest. The probability density function for $\mathbf{y}$ is

$$f(\mathbf{y}) = \int f(\mathbf{y} \mid \theta) f(\theta) d\theta.$$

The value, V, of the evidence (7.7) may then be rewritten in the following way

$$\frac{\int f(\mathbf{y} \mid \theta) f(\mathbf{x} \mid \theta) f(\theta) d\theta}{\int f(\mathbf{x} \mid \theta) f(\theta) d\theta \int f(\mathbf{y} \mid \theta) f(\theta) d\theta}. \tag{7.8}$$

For those unfamiliar with these kinds of manipulations, Bayes' theorem applied to conditional probability distributions is used to write $f(\theta \mid \mathbf{x})$ as $f(\mathbf{x} \mid \theta) f(\theta)/f(\mathbf{x})$. The law of total probability with integration replacing summation is used to write $f(\mathbf{x})$ as $\int f(\mathbf{x} \mid \theta) f(\theta)\, d\theta$.

7.3.1 Assessment of Similarity Only

The case in which a window is broken, a large group of fragments is found on the suspect and the group is similar in properties to the broken window was discussed earlier in Section 6.2.2 and the likelihood ratio is given by (6.25). The factor $1/\gamma_1$ in (6.25) is an approximation to the ratio given in (7.7).

An approach to the evaluation of V in (7.7) that only considers similarity is to consider the summary statistics for the recovered and control data. The available data $\mathbf{x}$ and $\mathbf{y}$ may be replaced by summary statistics for the means and variances, with

$$\bar{x} = \sum_{i=1}^{m} x_i/m \text{ and } \bar{y} = \sum_{j=1}^{n} y_j/n,$$

and

$$s_x^2 = \sum_{i=1}^{m} (x_i - \bar{x})^2/(m-1) \text{ and}$$

$$s_y^2 = \sum_{j=1}^{n} (y_j - \bar{y})^2/(n-1).$$

Then, following the argument of Walsh et al. (1996), the ratio (7.7) may be written as

$$V = \frac{f(\bar{x} - \bar{y} \mid \bar{x}, s_x, s_y, H_p)}{f(\bar{y} \mid \bar{x}, s_x, s_y, H_d)}. \tag{7.9}$$

The numerator of (7.9) may be taken to be a Student's t-density with a Welch modification (Welch, 1937) (Section A.3.4) when the data $\mathbf{x}$ and $\mathbf{y}$ are Normally distributed and the population variances σ_x^2 and σ_y^2, of which s_x^2 and s_y^2 are estimates, are not assumed equal. The t-statistic for the numerator, which is to be referred to the t-density with the Welch modification, is

$$t_W = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}}. \tag{7.10}$$

The statistic t_W does not have a t-distribution but may be approximated by a t-distribution with ν degrees of freedom, where ν may be estimated from the data as

$$\nu = \frac{\left(\frac{s_x^2}{m} + \frac{s_y^2}{n}\right)^2}{\left(\frac{s_x^4}{m^2(m-1)} + \frac{s_y^4}{n^2(n-1)}\right)}, \tag{7.11}$$

which need not necessarily be an integer. Density values for t_W are provided on many readily available statistical software packages (such as R).

The denominator of (7.9) is the value of the probability density for the relevant population of glass at $\bar{y}$. This is usually obtained from a kernel density estimate (see (7.22), Section 7.5.1).

Consider as an example the data in Table 7.2 with $\bar{y} = 1.519\,507\,3$, $\bar{x} = 1.519\,573\,0$, $s_y = 5.24 \times$

Table 7.2 Refractive indices of glass fragments for Johnston, recovered, and a control set with means, separate and pooled standard deviations (s.d.).

	Johnston	Control
	1.519 40	1.519 50
	1.519 46	1.519 52
	1.519 47	1.519 53
	1.519 48	1.519 56
	1.519 50	1.519 57
	1.519 52	1.519 59
	1.519 52	1.519 60
	1.519 53	1.519 60
	1.519 56	1.519 62
	1.519 57	1.519 64
	1.519 57	
Mean	1.519 507 3	1.519 573 0
S.d.	5.24×10^{-5}	4.55×10^{-5}
Pooled s.d.	4.92×10^{-5}	

The number of recovered fragments $n = 11$ and the number of control fragments $m = 10$.
Source: From Walsh et al. (1996). Reprinted with permissions of Elsevier.

10^{-5}, $s_x = 4.55 \times 10^{-5}$, $m = 10$, and $n = 11$. The value of the t-statistic in (7.10) using a pooled standard deviation, with 19 degrees of freedom, is 3.06. The 99.5% point of a t-distribution with 19 degrees of freedom is 2.86, so the null hypothesis that the recovered (Johnston) and control data are samples from populations with the same mean is rejected in favour of the two-sided alternative of samples from populations with different means at the 1% level. The significance probability (Section 3.6) for the two-sided test is 0.0064. Thus, the conclusion of a scientist who used this approach would be to reject the hypothesis that the glass fragments found on Johnston came from the crime scene window and this evidence would be discarded. However, one of the problems associated with the use of significance probabilities is that there is a dichotomy between data for which the null hypothesis is rejected and data for which the null hypothesis is not rejected, the 'fall-off-the-cliff' effect of Section 1.3.3.

The value of the numerator is obtained as the probability density at x equal to the Johnston mean 1.519 507 3 of a non-central t-distribution (Section A.3.4) with $\nu = 31.29$ degrees of freedom (derived from (7.11)), centred at the control mean, $\mu = 1.519\ 573\ 0$, with spread parameter λ equal to $\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$, which equals 2.14×10^{-5} in this example. Transformation to $y = (x - \mu)/\lambda$ gives a value for the t-statistic of 3.07 with 31.29 degrees of freedom. Reference to appropriate statistical

software (such as R) gives a value for the central t-density function of 0.006. An adjustment by the factor $1/\lambda$ gives a value for the non-central t-density of $0.006/2.14 \times 10^{-5}$ or 280. This is the value for the numerator.

The value of the denominator is obtained from population data and a kernel density estimate. For this example, the value of the density estimate at the Johnston mean is taken to be 109 (Walsh et al., 1996). The likelihood ratio is then $280/109 = 2.6$. This provides slight support for the proposition that the fragments found on Johnston's clothing come from the crime scene window. This conclusion is in contrast to the rejection of this proposition at the 1% level using a two-tailed test.

Goldmann et al. (2004) described another application of the t-distribution and the Welch modification, in which the determination of the source of illicit pills through examination of the dye present in the pills is assisted. The dye considered is CI 147 20. There is a sample Y with 5 pills to be compared with a specific batch X with 20 pills and with another batch containing 100 pills attributed to the same producer, Z, as the producer of X. The measurement of interest is the concentration of the dye, expressed as a percentage. Two pairs of propositions are compared. The first is as follows:

H_{p1}: sample Y comes from batch X;

H_{d1}: sample Y does not come from batch X.

The second is as follows:

H_{p2}: sample Y comes from a batch produced by Z;

H_{d2}: sample Y does not come from a batch produced by Z.

The summary statistics (Goldmann et al., 2004) are presented in Table 7.3.

No sample size is given for the general population. The population is characterised by illicit pills coloured with dye CI 14 720. The percentage concentration of the dye in the pills is Normally distributed with mean 0.300% and standard deviation 0.06%. This contrasts with the denominator in the previous example in which a kernel density estimate was used.

The degrees of freedom (ν) and pooled standard deviations (s_p) of Y with X and of Y with Z are $\nu_{y,x} = 4.50$, $s_{p,y,x} = 0.0092$, and $\nu_{y,z} = 5.75$, $s_{p,y,z} = 0.0098$.

Table 7.3 Summary statistics for concentration of dye CI 14 720 in illicit pills.

Sample	Sample Y	Batch X	Producer Z	General population
Size	5	20	100	—
Mean (%)	0.165	0.140	0.180	0.300
Standard deviation (%)	0.02	0.01	0.04	0.06

Source: From Goldmann et al. (2004). Reprinted with permissions of ASTM International.

Consider the first pair of propositions. The value of the evidence is obtained by comparing the probability density of $\bar{y}$ in batch X and the probability density of $\bar{y}$ in the general population. This is 5.4.

Consider the second pair of propositions. The value of the evidence is obtained by comparing the probability density of $\bar{y}$ in batches produced by Z and the probability density of $\bar{y}$ in the general population. This is 23.3.

7.3.2 Sources of Variation: Two-Level Models

Notice that there are often two sources of variation to be considered in the measurements. There is variation within a particular source and there is variation between sources.

For example, consider evidence of fragments of glass from a broken window from which refractive index (r.i.) measurements have been made. There is variation in the r.i. measurements amongst the different fragments of glass. These different measurements may be thought of as a sample from the population corresponding to all possible r.i. measurements from that particular window. The population has a mean, θ, say, and a variance σ^2. The measurements of r.i. of fragments from that window can be assumed to be Normally distributed with mean θ and variance σ^2. Secondly, there is variation in the r.i. mean θ between different windows. The mean θ has a probability

distribution with its own mean μ, say, and variance τ^2. Typically τ^2 will be much greater than σ^2. A Normal distribution is assumed for θ. However, a look at Figure 7.1, which is a histogram of r.i. measurements from 2269 examples of float glass from buildings given in Table 7.11 (Lambert and Evett, 1984), shows that this is not a particularly realistic assumption. A more realistic approach will be described in Section 7.5.2.

Similar considerations apply for other types of evidence. For measurements on the medullary widths of cat hairs, for example, there will be variation amongst hairs from the same cat and

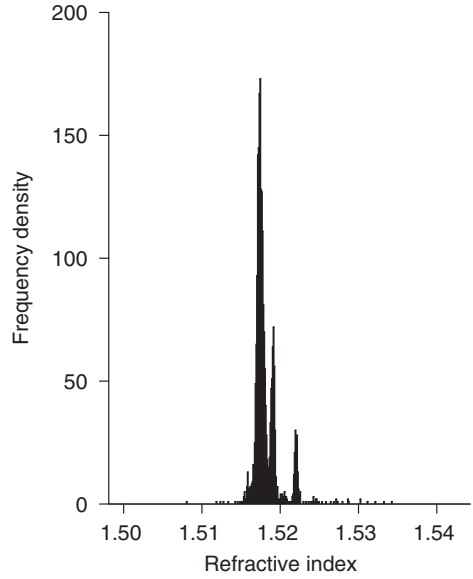

Figure 7.1 Refractive index measurements from 2269 fragments of float glass from buildings (Source: Modified from Lambert and Evett, 1984).

amongst hairs from different cats. For measurements on footprints there is considerable variability between footprints from different people and considerable similarity for multiple impressions taken from the same person (Kennedy et al., 2003). For handwriting evidence, there will be variation in characteristics between characters of a given type (say, letter *o*) from the same writer (the so-called *within-writer variability*) and from different writers (the so-called *between-writers variability*) (Marquis et al., 2005).

Thus, when considering the assessment of continuous data at least two sources of variability have to be considered: the variability within the source (e.g. window or writer) from which the measurements were made and the variability between the different possible sources (e.g. windows or writers). In some cases, a third level of variability must be added alongside the levels described earlier. For glass evidence, for example, one may need to model the measurement error (i.e. the error due to the precision of the instrument) separately. Three level models will be addressed in Section 7.6.5. Such models are known as hierarchical.

7.3.3 Transfer Probability

Consider transfer of material from the crime scene to the criminal. A PoI is found with similar material on his clothing, say. This material may have come from the crime scene. Alternatively, it may have come from somewhere else under perfectly

innocent circumstances. There are two sets of circumstances to consider. First, conditional on the PoI having been present at the crime scene (H_p), there is a probability that material will have been transferred from the scene to the PoI. It has also to be borne in mind that someone connected with the crime may have had no fragments transferred from the scene to their person and have had fragments similar to those found at the crime scene transferred to their person from somewhere else by innocent means. Secondly, there is the probability that a person unconnected with the crime (i.e. this is conditional on a PoI not having been present at the scene, H_d) will have material similar to the crime material on their person.

Consider the case of glass fragments as described by Evett (1986). Let t_i $(i = 0, 1, 2, \ldots)$ be the probability that, given H_p, i fragments of glass would have been transferred. More correctly, let t_n be the probability that, given H_p, the presence of the PoI at the crime scene, n fragments would be found on the clothing of the PoI on searching. This allows not only for the mechanism of transfer but also for the mechanisms of persistence and recovery. Let $b_{1,m}$ $(m = 1, 2, \ldots)$ denote the probability of the chance occurrence of a single group of m fragments of glass on their clothing (see Section 5.3.2.1). In general, the probability of the chance occurrence of g groups, with $m_1, \ldots, m_g$ fragments of glass from each group, on their clothing can be denoted $b_{g,m_1,\ldots,m_g}$ or $\{b_{g,\mathbf{m}}; \mathbf{m} = (m_1, \ldots, m_g)^T\}$.

Also, $b_{1,s}$ and $b_{1,l}$ denote the conditional probabilities that, if one group of fragments is found, it contains a small (s) or large (l) number of fragments, as described in Section 5.5.3.

Consider the case where a single fragment has been found. A general expression for more groups of fragments is given in Evett (1986).

The evidence E consists of three parts. The first part is the existence (m_1) of one fragment on the clothing of the person of interest. The second part is that its refractive index is y. This is the transferred particle form of the evidence. The probability that a person in the relevant population has a fragment of glass on their clothing may be denoted $\Pr(m_1 \mid H_p)$ or $b_{1,1}$. The probability that a person in the general population does not have a fragment of glass on their clothing is denoted b_0. The third part of the evidence are the measurements $\mathbf{x}$ on the control material. This is relevant for the determination of the numerator but not the denominator.

Consider the denominator of the likelihood ratio. This is

$$\begin{aligned}\Pr(E \mid H_d, I) &= \Pr(m_1, y \mid H_d, I) \\ &= \Pr(m_1 \mid H_d, I) \times f(y \mid H_d, I, m_1) \\ &= b_{1,1} f(y \mid H_d, I, m_1).\end{aligned}$$

The probability density function $f(y \mid H_d, I, m_1)$ is taken to be a Normal density function, with mean μ and variance τ^2 (or, more correctly, $\tau^2 + \sigma^2$) as is illustrated in Section 7.4.2.

Consider the numerator of the likelihood ratio. If the PoI was present at the crime scene, there are two possible explanations for the presence of the glass fragment on the clothing of the PoI. Either the fragment has been acquired by innocent means (an event with probability $b_{1,1}$) and no fragment has been transferred from the crime scene (an event with probability t_0), or the fragment was transferred from the crime scene and none was transferred by innocent means, two events with probabilities t_1 and b_0, respectively (these explanations can be relaxed to take into account different situations such as the presence of two groups of recovered material, one transferred from the crime scene, the other transferred by innocent reasons, as in Section 6.2.2). Let $\mathbf{x}$ denote the measurements on the source (control) sample. The numerator is then

$$t_0 b_{1,1} f(y \mid H_d, I, m_1) + t_1 b_0 f(y \mid H_p, \mathbf{x}, I),$$

where $f(y \mid H_p, \mathbf{x}, I)$ is taken to be Normal with mean $\bar{x}$ and variance σ^2. Notice the terms $t_0 b_{1,1}$ and $t_1 b_0$. The former is the probability that no particle is transferred from the crime scene and one particle is transferred from the background. The latter is the probability that one particle is transferred from the crime scene and no particle is transferred from the background. In the term involving H_d the fragment is assumed to have been transferred by innocent means (see Section 5.3.2.2). The probability density function for y

in this situation is then the one that holds when the PoI is unconnected with the crime. Hence, the conditioning on H_d is permissible.

The likelihood ratio is then

$$V = t_0 + \frac{t_1 b_0}{b_{1,1}} \frac{f(y \mid H_p, \mathbf{x}, I)}{f(y \mid H_d, m_1, I)}. \qquad (7.12)$$

This is the equivalent of (5.7) in Section 5.3.2.2 where $1/\gamma = f(y \mid H_p, \mathbf{x}, I)/f(y \mid H_d, m_1, I)$.

Another derivation for continuous data is given by (7.18) in Section 7.4.3.

7.4 NORMAL BETWEEN-SOURCE VARIATION

The approach to evidence evaluation described in Section 7.3 was first proposed by Lindley (1977c) in the context of a problem involving the measurements of the refractive index of glass. These measurements may be made on fragments of glass at the scene of a crime and on fragments of window glass found on the clothing of a PoI (see Example 1.2 in Section 1.3.3). These measurements are subject to error and it is this which is incorporated into V.

7.4.1 Marginal Distribution of Measurements

Let x be a measurement from a particular control fragment. Let the mean of measurements

from the source of this fragment be θ_1. Let y be one measurement from a particular recovered fragment. Let the mean of measurements from the source of this fragment be θ_2. The variance of measurements within a source is assumed constant amongst sources and is denoted σ^2. The dependence of the distribution of these measurements on the control source from which they come can be made explicit in the notation. The distributions of X and Y, given θ_1, θ_2, and σ^2, are

$$(X \mid \theta_1, \sigma^2) \sim N(\theta_1, \sigma^2),$$
$$(Y \mid \theta_2, \sigma^2) \sim N(\theta_2, \sigma^2),$$

where the dependence on θ_1 or θ_2 and σ^2 is made explicit. Notice, also, that variation in X is modelled. Contrast this approach with the coincidence probability approach of Section 3.7 in which the mean of measurements on the control fragments was taken as fixed. The conditioning on H_d is implicit. The means θ_1 and θ_2 of these distributions may themselves be thought of as observations from another distribution (that of variation between sources), which for the present is taken to be Normal, with mean μ and variance τ^2. Thus, θ_1 and θ_2 have the same probability density function and

$$(\theta \mid \mu, \tau^2) \sim N(\mu, \tau^2).$$

The distributions of X and of Y, independent of θ, can be determined by taking the so-called

convolutions of x and of y with θ to give

$$f(x \mid \mu, \sigma^2, \tau^2) = \int f(x \mid \theta, \sigma^2) f(\theta \mid \mu, \tau^2) d\theta$$
$$= \int \frac{1}{2\pi\sigma^2\tau^2} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}$$
$$\exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta$$
$$= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \exp\left\{-\frac{1}{2(\sigma^2+\tau^2)}(x-\mu)^2\right\},$$

using the result that

$$\frac{1}{2\sigma^2}(x-\theta)^2 + \frac{1}{2\tau^2}(\theta-\mu)^2 = \frac{(\theta-\mu_1)^2}{\tau_1^2} + \frac{(x-\mu)^2}{\sigma^2+\tau^2},$$

where

$$\mu_1 = \frac{\sigma^2\mu + \tau^2 x}{\sigma^2+\tau^2},$$
$$\tau_1^2 = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}.$$

Similarly

$$f(y \mid \mu, \sigma^2, \tau^2) = \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \exp\left\{-\frac{1}{2(\sigma^2+\tau^2)}(y-\mu)^2\right\}.$$

Notice that τ^2 has been omitted from the distributions of x and y, given θ_1, θ_2, and σ^2. This is

because the distributions of x and y, given these parameters, are independent of τ^2. Similarly, the distribution of θ, given μ and τ^2, is independent of σ^2.

The effect of the two sources of variability is that the mean of the r.i. measurements is the overall mean μ and the variance is the sum of the two component variances σ^2 and τ^2. The distribution remains Normal. Thus

$$(X \mid \mu, \sigma^2, \tau^2) \sim N(\mu, \sigma^2 + \tau^2),$$
$$(Y \mid \mu, \sigma^2, \tau^2) \sim N(\mu, \sigma^2 + \tau^2). \qquad (7.13)$$

7.4.2 Approximate Derivation of the Likelihood Ratio

Consider an application to a broken window as in Example 1.2 (Section 1.3.3). A crime is committed in which a window is broken. A PoI is apprehended soon afterwards and a fragment of glass is found on their clothing. Its refractive index is y. A sample of m fragments is taken from the broken window at the scene of the crime and their refractive index measurements are $\mathbf{x} = (x_1, \ldots, x_m)^T$, with mean $\bar{x}$. The two propositions to be compared are as follows:

H_p, the recovered fragment is from the crime scene window;

H_d, the recovered fragment is not from the crime scene window.

An approximate derivation of the likelihood ratio may be obtained by replacing θ by $\bar{x}$ in the distribution of y so that $f(y \mid \theta, \sigma^2)$ becomes $f(y \mid \bar{x}, \sigma^2)$. (See (4.32) for a similar result using a uniform prior for a Normal distribution.) This is only an approximate distributional result. A more accurate result is given later to account for the sampling variability of $\bar{x}$. For the present, an approximate result for the numerator is that

$$(Y \mid \bar{x}, \sigma^2, H_p, I) \sim N(\bar{x}, \sigma^2), \tag{7.14}$$

an application of (4.32). Also, from (7.13)

$$(Y \mid \mu, \sigma^2, \tau^2, H_d, I) \sim N(\mu, \tau^2 + \sigma^2). \tag{7.15}$$

For τ^2 much greater than σ^2, assume also that $\tau^2 + \sigma^2$ can be approximated by τ^2. The likelihood ratio is then

$$\begin{aligned} V &= \left[\frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{(y-\bar{x})^2}{2\sigma^2}\right\}\right] \Bigg/ \\ &\qquad \left[\frac{1}{\tau\sqrt{(2\pi)}} \exp\left\{-\frac{(y-\mu)^2}{2\tau^2}\right\}\right] \\ &= \frac{\tau}{\sigma} \exp\left\{\frac{(y-\mu)^2}{2\tau^2} - \frac{(y-\bar{x})^2}{2\sigma^2}\right\} \end{aligned}$$

(Evett, 1986). Note that this likelihood ratio depends on an assumption that fragments from a single source are found on the PoI and that these have come from the crime scene.

This result has some intuitively attractive features. The likelihood ratio is larger for values of y, which are further from μ and are therefore assumed to be rarer; that is, the rarer the value of the refractive index of the recovered fragment, the larger the likelihood ratio. Also, the larger the value of $|\, y - \bar{x} \,|$, the smaller the value of the likelihood ratio; that is, the further the value of the refractive index of the receptor glass fragment is from the mean of the values of the r.i's of the source fragments, the smaller the likelihood ratio.

Values for τ equal to 4×10^{-3} and for σ equal to 4×10^{-5} are given by Evett (1986). Values of V for various values of $(y - \mu)/\tau$ and of $(y - \bar{x})/\sigma$, the standardised distances of y from the overall mean and the source mean, are given in Table 7.4. Note that the ratio $\tau/\sigma = 100$, giving ample justification for the approximation τ^2 to the variance of y, given

Table 7.4 Likelihood ratio values for varying values of $(y - \bar{x})/\sigma$ and $(y - \mu)/\tau$.

$(y - \bar{x})/\sigma$	$(y - \mu)/\tau$		
	0	1	2
0	100	165	739
1	61	100	448
2	14	22	100
3	1	2	8

H_d, earlier. Also, this ratio is a large contributor to the value of V.

7.4.3 Lindley's Approach

A more detailed analysis was provided by Lindley (1977c). Assume, as before, that the measurements are distributed about the true unknown value, θ, of the refractive index with a Normal distribution and a known variance σ^2 and that the propositions H_p and H_d to be compared are as in Section 7.4.2. If m measurements are made at the scene (source measurements, $x_1, \ldots, x_m$), then it is sufficient to consider the sample mean, $\bar{x} = \sum_{i=1}^{m} x_i/m$. Conditional on θ_1, the mean of the r.i. measurements of the crime window, the sample mean $\bar{X}$ is Normally distributed about θ_1 with variance σ^2/m (Section A.3.2). Let $\bar{y}$ denote the sample mean of n similar measurements (recovered measurements, $y_1, \ldots, y_n$) made on material found on the PoI; conditional on θ_2, $\bar{Y}$ is Normally distributed about θ_2 with variance σ^2/n. In the case H_p holds, where the source and recovered measurements come from the same source, $\theta_1 = \theta_2$. Otherwise, in the case H_d holds, $\theta_1 \neq \theta_2$.

The distribution of the true unknown value θ has also to be considered. There is considerable evidence about the distribution of r.i.s.; see, for example, Curran et al. (2000). First, assume as before that the true unknown values θ are Normally distributed about a mean μ with variance

τ^2, both of which are assumed known. Typically τ will be larger, sometimes much larger, than σ (see earlier, where $\tau/\sigma = 100$). This assumption of Normality is not a realistic one in this context where the distribution has a pronounced peak and a long tail to the right; see Figure 7.1. However, the use of the Normality assumption enables analytic results to be obtained as an illustration of the general application of the method. The unconditional distributions of the means of the control and recovered measurements (with m measurements on the control and n on the recovered sample), $\bar{X}$ and $\bar{Y}$, in the denominator are independent and are, respectively, $N(\mu, \tau^2 + \sigma^2/m)$ and $N(\mu, \tau^2 + \sigma^2/n)$.

Let $\sigma_1^2 = \tau^2 + \sigma^2/m$ and $\sigma_2^2 = \tau^2 + \sigma^2/n$ where τ^2 is the between-source variance. Then $(\bar{X} - \bar{Y}) \sim N(0, \sigma_1^2 + \sigma_2^2)$ and $Z = (\sigma_2^2\bar{X} + \sigma_1^2\bar{Y})/(\sigma_1^2 + \sigma_2^2)$ is distributed as $N(\mu, \sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2))$, and $(\bar{X} - \bar{Y})$, and Z are also independent. The denominator may then be written as

$$\frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{(\bar{x} - \bar{y})^2}{2(\sigma_1^2 + \sigma_2^2)}\right\}$$
$$\exp\left\{-\frac{(z - \mu)^2(\sigma_1^2 + \sigma_2^2)}{2\sigma_1^2\sigma_2^2}\right\}.$$

In the numerator, it can be shown that the joint unconditional distribution of $\bar{X}$ and $\bar{Y}$ is bivariate Normal with means μ, variances σ_1^2 and σ_2^2, and covariance τ^2 (Section A.3.9). The

distribution of $\bar{X} - \bar{Y}$ is $N\left(0, \sigma^2(\frac{1}{m} + \frac{1}{n})\right)$. Let $W = (m\bar{X} + n\bar{Y})/(m+n)$. The distribution of W is $N(\mu, \tau^2 + \sigma^2/(m+n))$. Also, $(\bar{X} - \bar{Y})$ and W are independent. Let $a^2 = 1/m + 1/n$ and $\sigma_3^2 = \tau^2 + \sigma^2/(m+n)$. Then the numerator may be written as

$$\frac{1}{2\pi a\sigma\sigma_3} \exp\left\{-\frac{(\bar{x}-\bar{y})^2}{2a^2\sigma^2}\right\} \exp\left\{-\frac{(w-\mu)^2}{2\sigma_3^2}\right\}.$$

The value, V, of the evidence is the ratio of the numerator to the denominator; after some simplification, this is

$$\frac{\sigma_1\sigma_2}{a\sigma\sigma_3} \exp\left\{-\frac{(\bar{x}-\bar{y})^2\tau^2}{a^2\sigma^2(\sigma_1^2+\sigma_2^2)}\right\} \exp\left\{-\frac{(w-\mu)^2}{2\sigma_3^2} + \frac{(z-\mu)^2(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2}\right\}.$$

Large values of this provide good evidence that the suspect was at the crime scene.

This expression may be simplified. Typically, τ is much larger than σ. Then $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \tau^2$, $Z = (\bar{X} + \bar{Y})/2$, and

$$V \simeq \frac{\tau}{a\sigma} \exp\left\{-\frac{(\bar{x}-\bar{y})^2}{2a^2\sigma^2}\right\} \exp\left\{-\frac{(w-\mu)^2}{2\tau^2} + \frac{(z-\mu)^2}{\tau^2}\right\}. \quad (7.16)$$

If the number of control measurements equals the number of recovered measurements then $m = n,\ z = w = \frac{1}{2}(\bar{x} + \bar{y})$, and

$$V \simeq \frac{m^{1/2}\tau}{2^{1/2}\sigma} \exp\left\{-\frac{m(\bar{x}-\bar{y})^2}{4\sigma^2}\right\} \exp\left\{\frac{(z-\mu)^2}{2\tau^2}\right\}. \tag{7.17}$$

Consider the particular case where $m = n = 1$. According to (7.17), V consists of two factors that depend on the measurements. The first is $\exp\{-(\bar{x}-\bar{y})^2/4\sigma^2\}$. This compares the absolute difference $|\ \bar{x}-\bar{y}\ |$ of the control and recovered measurements with their standard deviation $\sigma\sqrt{2}$ on the proposition $(\theta_1 = \theta_2)$ that they come from the same source. Let

$$|\ \bar{x}-\bar{y}\ |\ /\sigma\sqrt{2} = \lambda.$$

Then the value of the first factor is $\exp(-\lambda^2/2)$. A large value of λ favours the hypothesis that the two fragments come from different sources. This factor has an effect like that of a significance test of a null hypothesis of identity $(\theta_1 = \theta_2)$.

The second factor, $\exp\{(z-\mu)^2/2\tau^2\}$, with $z = \frac{1}{2}(\bar{x} + \bar{y})$ measures the typicality of the two measurements. This factor takes its smallest value, 1, when $z = \mu$ and increases as $|\ z-\mu\ |$ increases relative to its standard deviation. Thus the more unusual the glass (i.e. the larger the value of $|\ z-\mu\ |$), the greater the value of V and the stronger the inference in favour of a common source for the two measurements. Consider the

comment by Parker and Holford (1968) in Section 3.8. The first factor considers similarity. The second factor considers typicality. The assessment of similarity is not by-passed.

Note again that result (7.17) assumes implicitly that the fragments are from a single source. Denote this assumption by S. Then the aforementioned result is the ratio of the probability density functions $f(\bar{x}, \bar{y} \mid H_p, S)/f(\bar{x}, \bar{y} \mid H_d, S)$. A result including S as one of the uncertain elements and deriving an expression for $f(\bar{x}, \bar{y}, S \mid H_p)/f(\bar{x}, \bar{y}, S \mid H_d)$ was given by Grove (1980). Let T denote the event that fragments were transferred from the broken window to the suspect and persisted there until discovery by the police. Let A be the event that the suspect came into contact with glass from some other source. Assume that $\Pr(A \mid H_p) = \Pr(A \mid H_d) = p_A$, that $\Pr(T \mid H_p) = p_T$, that A and T are independent given H_p and that $\Pr(T \mid H_d) \simeq 0$. Grove (1980) shows that

$$\begin{aligned} V &= \frac{f(\bar{x}, \bar{y}, S \mid H_p)}{f(\bar{x}, \bar{y}, S \mid H_d)} \\ &= 1 + p_T \left\{ (p_A^{-1} - 1) \frac{f(\bar{x}, \bar{y} \mid H_p)}{f(\bar{x}, \bar{y} \mid H_d)} - 1 \right\} \\ &= (1 - p_T) + \frac{p_T(1 - p_A)}{p_A} \times \frac{f(\bar{x}, \bar{y} \mid H_p)}{f(\bar{x}, \bar{y} \mid H_d)}, \end{aligned} \quad (7.18)$$

where $f(\bar{x}, \bar{y} \mid H_p)/f(\bar{x}, \bar{y} \mid H_d)$ is the ratio of Lindley (1977c). The value derived by Grove (1980) takes account of transfer and persistence in a

way already derived for discrete data (see Section 6.2.3).

7.4.4 Interpretation of Result

Compare the result in (7.18) with the alternative derivation for the likelihood ratio obtained in (7.12) where t_0 replaces $(1 - p_T)$, t_1 replaces p_T, b_0 replaces $(1 - p_A)$, and $b_{1,1}$ replaces p_A. The ratio of the density functions $f(y \mid H_p, \mathbf{x}, I)$ and $f(y \mid H_d, m_1, I)$ was considered earlier. The extension described in (7.12) accounts for possible different sources of the fragment. For the single fragment case

$$V = t_0 + \frac{t_1 b_0}{b_{1,1}} \frac{\sqrt{\sigma^2 + \tau^2}}{\sigma} \exp\left\{ \frac{(y-\mu)^2}{2(\tau^2+\sigma^2)} - \frac{(y-\bar{x})^2}{2\sigma^2} \right\}.$$

Illustrative values for the distributional parameters and for the transfer probabilities from Evett (1986) are given in Tables 7.5 and 7.6.

The values for $b_0, b_{1,1}$ are suggested by Evett (1986) who cited Pearson et al. (1971). Evett

Table 7.5 Distributional parameters for glass problems.

μ	τ	σ
1.5186	4×10^{-3}	4×10^{-5}

Table 7.6 Transfer probabilities for glass problems.

b_0	$b_{1,1}$	t_0	t_1
0.37	0.24	0	0.056

contrasted the probabilities in Table 7.6 with results from Harrison et al. (1985) in which the numbers of people with one and two fragments on their clothing were proportionately closer. There is a need for a closer investigation of the estimation of these probabilities. Further examples are given in Section 7.5.2. These transfer probabilities are provided by Evett (1986) citing a personal communication by C.F. Candy. These probabilities are different from those given in Section 6.2.2. No claim is made that either set is definitive. All these probabilities are provided primarily for illustrative purposes. In practice more up-to-date values should be used. It is the responsibility of the expert to ensure the values used are of relevance.

From these values $t_1 b_0 / b_{1,1} = 0.086$. Notice that $\tau/\sigma = 100$ and that $\tau^2 + \sigma^2 \simeq \tau^2$. Thus

$$\frac{\sqrt{\tau^2 + \sigma^2}}{\sigma} \simeq \frac{\tau}{\sigma}.$$

For the single fragment case

$$V \simeq 8.6 \exp\left\{ \frac{(y-\mu)^2}{2\tau^2} - \frac{(y-\bar{x})^2}{2\sigma^2} \right\}. \qquad (7.19)$$

Table 7.7 Some values for the likelihood ratio V for the single fragment case.

$(y-\bar{x})/\sigma$	$(y-\mu)/\tau$		
	0.0	1.0	2.0
0.0	9	14	63
1.0	5	9	38
2.0	1	2	9
3.0	0.1	0.2	0.7

Source: Evett (1986). Reproduced with permission of Elsevier.

Some values for varying values of $(y-\bar{x})/\sigma$ and $(y-\mu)/\tau$ (standardised differences of y from the sample mean of the source fragments and from the overall mean) are given in Table 7.7. Small values of $(y-\bar{x})/\sigma$ imply similarity between source and recovered fragments. Small values of $(y-\mu)/\tau$ imply a common value of y. Notice the largest value of V is given by a small value of $(y-\bar{x})/\sigma$ and a large value of $(y-\mu)/\tau$.

7.4.5 Examples

Consider a case with only one control ($m=1$) and one recovered ($n=1$) fragment of glass. Denote these as $\bar{x}$ and $\bar{y}$, for consistency with (7.17). Suppose the ratio of between-groups standard deviation τ to within-groups standard deviation σ is 100. The control and recovered measurements are found to be two within-group

standard deviations apart. Since the variance of $(\bar{x} - \bar{y})$ equals $2\sigma^2$, this separation implies that $|\bar{x} - \bar{y}| / \sigma\sqrt{2} = 2$. A conventional significance test (Section 3.6) would reject the hypothesis of a common source at the 5% level of significance. Assume that the mid-point of $\bar{x}$ and $\bar{y}$, which is $(\bar{x} + \bar{y})/2$, the mean denoted z in (7.17) is the population mean μ. Then, from (7.17)

$$V = \frac{100e^{-2}}{\sqrt{2}} = 9.57.$$

The odds in favour of a common source are increased by a factor of almost 10, a result in contrast to the rejection at the 5% level of significance in a conventional significance test. The values of (7.17) for $\tau/\sigma = 100$ as a function of λ and $\delta = |z - \mu| / \tau$, the deviation of the mean of the two measurements from μ, standardised on the assumption that the hypothesis of a common source is true are given in Table 7.8. For other values of τ/σ, multiply entries by $\tau/(100\sigma)$.

Consider the more general formula for V, as given in (7.16) in Section 7.4.3, namely,

$$V \simeq \frac{\tau}{a\sigma} \times \exp\left\{-\frac{(\bar{x} - \bar{y})^2}{2a^2\sigma^2}\right\} \times \exp\left\{-\frac{(w - \mu)^2}{2\tau^2} + \frac{(z - \mu)^2}{\tau^2}\right\}. \quad (7.20)$$

The following information is needed in order that V may be evaluated:

Table 7.8 Value of $\tau(2^{1/2}\sigma)^{-1}\exp(-\frac{1}{2}\lambda^2 + \frac{1}{2}\delta^2)$ (7.17) as a function of $\lambda = |\bar{x} - \bar{y}|/(2^{1/2}\sigma)$ and $\delta = |z - \mu|/\tau$ for $\tau/\sigma = 100$.

δ	λ				
	0	1.0	2.0	4.0	6.0
0	70.7	42.9	9.57	0.024	1.08×10^{-6}
1.0	117	70.7	15.8	0.039	1.78×10^{-6}
2.0	522	317	70.7	0.175	7.94×10^{-6}
3.0	6370	3860	861	2.14	9.71×10^{-5}

- The number of control measurements (m);
- The mean of the control measurements ($\bar{x}$);
- The number of recovered measurements (n);
- The mean of the recovered measurements ($\bar{y}$);
- The variance (assumed known) of the measurements on the control and recovered samples (σ^2);
- The overall mean (assumed known) of the refractive indices (μ);
- The overall variance (assumed known) of the refractive indices (τ^2).

The following values may be derived from the earlier:

- $z = (\bar{x} + \bar{y})/2$;
- $w = (m\bar{x} + n\bar{y})/(m + n)$;
- $a^2 = 1/m + 1/n$.

In the following numerical example using data from Evett (1977) and Lindley (1977c), we have $\bar{x} = 1.518\,458$, $m = 10$; $\bar{y} = 1.518\,472$, $n = 5$; $\sigma = 0.000\,04$; $\tau = 0.004$. The overall mean μ is taken to be 1.518 2 and has been derived from the 2 269 measurements for building float glass published by Lambert and Evett (1984); see Figure 7.1. With these figures, $a^2 = 0.3, w = 1.518\,463, z = 1.518\,465$, and

$$\frac{\tau}{a\sigma} = 182.5742,$$

$$\frac{(\bar{x} - \bar{y})^2}{2a^2\sigma^2} = 0.2042,$$

$$\frac{(w - \mu)^2}{2\tau^2} = 0.002\,16,$$

$$\frac{(z - \mu)^2}{\tau^2} = 0.004\,39,$$

$$V = 149.19.$$

The odds in favour of the suspect being at the crime scene are thus increased by a factor of 150.

7.5 NON-NORMAL BETWEEN-SOURCE VARIATION

The normality of observations often represents a convenient assumption to model between-source variation (as in Section 7.4). However, it must be acknowledged that there are many practical situations where available measurements do not have regular characteristics that make it suitable

to use such standard parametric models. In particular, not all data are unimodal, symmetric, and bell-shaped and may not be modelled by a Normal distribution. The histogram of the refractive index of glass fragments (Figure 7.1) and a histogram of the medullary width of cat hairs (Figure 7.2 from data in Table 7.9) both illustrate this. In such cases, an estimation of the probability density function may provide a better representation of available data.

7.5.1 Estimation of a Probability Density Function

The estimation of a population mean (μ) and population variance (σ^2) by a sample mean ($\bar{x}$) and variance (s^2) of data sampled from a relevant

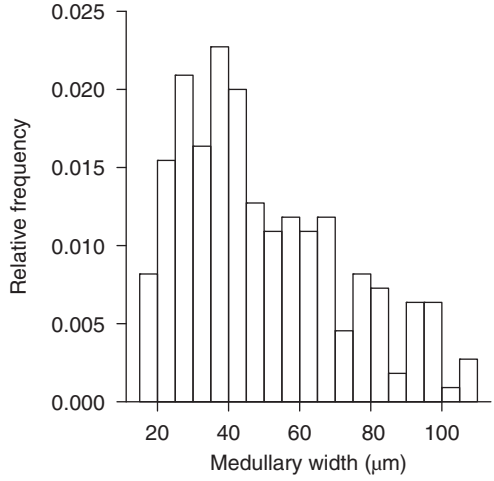

Figure 7.2 Medullary width (in microns) of 220 cat hairs (Source: Based on Peabody et al., 1983).

Table 7.9 Medullary widths in microns of 220 cat hairs (Source: Based on Peabody et al., 1983).

17.767	28.600	39.433	52.233	68.467
18.633	28.600	39.867	52.867	69.333
19.067	29.033	39.867	53.300	71.067
19.067	29.033	39.867	53.300	71.500
19.067	29.467	40.300	53.733	71.667
19.133	30.333	40.300	53.733	73.233
19.300	30.767	40.733	54.167	74.533
19.933	31.200	41.167	54.600	75.400
19.933	31.200	41.167	55.033	76.267
20.367	31.300	41.600	55.467	76.267
20.367	31.633	41.600	55.900	77.133
20.367	31.633	42.033	56.767	77.367
20.600	31.807	42.467	57.200	78.000
20.800	32.000	42.467	57.200	78.000
20.800	32.067	42.467	57.200	79.500
21.233	32.067	42.467	57.633	79.733
21.233	32.500	42.467	58.067	80.167
21.400	32.500	42.467	58.067	80.167
22.533	33.800	42.900	58.500	80.167
22.967	33.800	42.900	58.933	81.467
22.967	34.233	42.900	58.933	81.467
23.400	34.667	42.900	60.233	81.900
23.833	34.667	43.333	60.533	82.767
23.833	35.533	44.200	60.667	84.067
24.267	35.533	44.200	60.667	87.100
24.700	35.533	44.300	60.667	87.967
25.133	35.533	45.067	61.100	90.133
25.133	35.533	45.933	61.967	90.267
25.133	36.400	45.933	62.400	91.867
25.133	36.400	45.933	62.400	91.867
25.300	37.267	46.150	63.000	92.733
26.000	37.267	46.583	63.267	93.167
26.000	37.267	46.800	63.700	93.600
26.233	38.567	46.800	65.433	95.333

(*continued*)

Table 7.9 (*Continued*)

26.433	38.567	47.167	65.867	96.267
26.433	38.567	48.100	66.300	97.067
26.867	39.000	48.317	66.733	97.500
26.867	39.000	48.967	66.733	97.500
27.133	39.000	48.967	66.733	97.933
27.733	39.000	49.400	67.167	99.667
27.733	39.000	50.267	67.600	100.100
27.733	39.433	51.567	67.600	106.600
28.167	39.433	51.567	68.033	106.600
28.167	39.433	52.000	68.033	107.467

population is a common idea. Also, the probability density function itself may be estimated from data taken from the population.

Estimation of a probability density function is not too difficult so long as the distribution is fairly smooth. A procedure known as *kernel density estimation* is used; see Silverman (1986) for technical details. For early applications to forensic science, see Aitken and MacDonald (1979) for an application with discrete data to forensic odontology and Aitken (1986) for an application to the discrimination between cat and dog hairs in which two variables are considered. An example is given here of the application of the technique to the distribution of the medullary width of cat hairs.

Consider data on the medullary widths (in microns) of 220 cat hairs (Peabody et al., 1983). A version of these modified to make the analysis

easier is given in Table 7.9 and a histogram to illustrate the distribution is shown in Figure 7.2 from which it can be seen the data are positively skewed and perhaps not unimodal. The histogram has been constructed from the full data set by selecting intervals of fixed width and fixed boundary points, namely, 15.01–20.00, 20.01–25.00, . . . , 105.01–110.00 microns. Individual observations are then allocated to the appropriate interval and a frequency count obtained. Each interval is five units (microns) wide and there are 220 observations. If each observation is allocated unit height the total area encompassed by the histogram is 5×220 equals 1100 units. Thus if the height of each bar of the histogram is reduced by a factor of 1100, the area under the new diagram is 1. This new histogram may be considered a very naïve probability function (with steps at the boundary points of the bars of the histogram).

The method of kernel density estimation may be considered as a development of the histogram. Consider the histogram to be constructed with rectangular blocks, each block corresponding to one observation. The block is positioned according to the interval in which the observation lies. The method of kernel density estimation used here replaces the rectangular block by a Normal probability density curve, known in this context as the *kernel function*. The curve is positioned by centring it over the observation to which it relates. The estimate of the probability density curve is

then obtained by adding the individual curves together over all the observations in the data set and then dividing this sum by the number of observations. Since each component of the sum is a probability density function, each component has area 1. Thus, the sum of the functions divided by the number of observations also has area 1 and is a probability density function.

In the construction of a histogram a decision has to be made initially as to the width of the intervals. If these are wide, the histogram is very uninformative regarding the underlying distribution. If these are narrow, there is too much detail and general features of the distribution are lost. Similarly, in kernel density estimation, the spread of the Normal density curves has to be determined. The spread of the curves is represented by the variance. If the variance is chosen to be large, the resultant estimated curve is very smooth. If the variance is chosen to be small, the resultant curve is very spikey (see Figure 7.3).

Mathematically, the kernel density estimate of an underlying probability density function can be constructed as follows. The discussion is in the context of estimating the distribution of the medullary widths of cat hairs. There is variation in the medullary width both within hairs from an individual cat and between different cats. Denote the measurement of the mean medullary width of hairs from a particular cat by θ. The corresponding probability density function $f(\theta)$ is to be estimated. A training data set $D = \{z_1, \ldots, z_k\}$ is

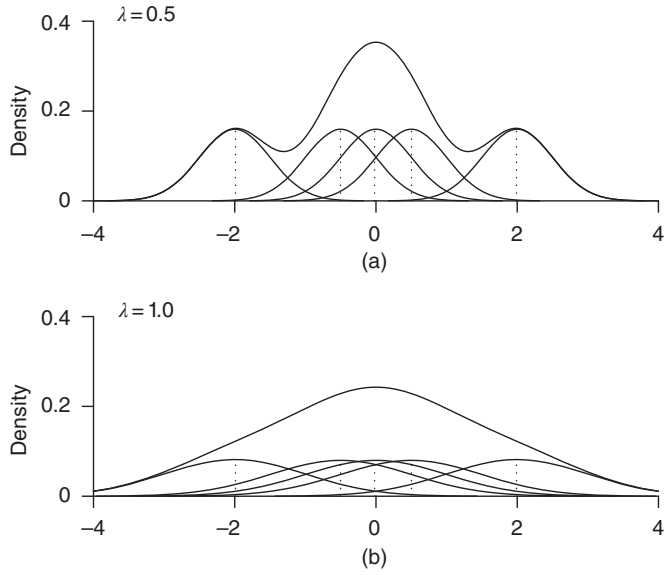

Figure 7.3 Examples of kernel density estimates showing individual kernels. Smoothing parameter values are (a) $\lambda = 0.5$ and (b) $\lambda = 1$.

available to enable this to be done. The variance of the width of hairs from different cats is estimated by

$$s^2 = \sum_{i=1}^{k} (z_i - \bar{z})^2/(k-1), \qquad (7.21)$$

where $\bar{z}$ is the sample mean. This variance is a mixture of the variances measuring the variability of the medullary width between and within cats and will be used as an approximation to the variance of the medullary width between cats. The sample standard deviation s is then multiplied by

a parameter, known as the *smoothing parameter*, denoted here by λ, which determines the smoothness of the density estimate. The kernel density function $K(\theta \mid z_i, \lambda)$ for point z_i is then taken to be a Normal distribution with mean z_i and variance $\lambda^2 s^2$,

$$K(\theta \mid z_i, \lambda) = \frac{1}{\lambda s\sqrt{2\pi}} \exp\left\{-\frac{(\theta - z_i)^2}{2\lambda^2 s^2}\right\}.$$

The estimate $\hat{f}(\theta \mid D, \lambda)$ of the probability density function is then given by

$$\hat{f}(\theta \mid D, \lambda) = \frac{1}{k}\sum_{i=1}^{k} K(\theta \mid z_i, \lambda). \qquad (7.22)$$

Notice here that there is an implicit assumption that a suitable data set D exists and that it is a data set from a relevant population. This latter comment is of particular relevance when considering DNA profiling where there is much debate as to the choice of the relevant population in a particular case. Also, if data were available on variability within groups, an adjustment can be made to the estimate of the between-group variance s^2 (7.21). Consider data of the form $\{z_{ij}, i = 1, \ldots, k, j = 1, \ldots l\}$ where k is the number of groups and l is the number of members of each group, assumed constant amongst groups. Let $\bar{z}_i$ denote the mean of the i-th group and $\bar{z}$ the overall mean. The within-group variance σ^2 is

then estimated by

$$\hat{\sigma}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} (z_{ij} - \bar{z}_i)^2 / (kl - k)$$

and the between-group variance τ^2 by

$$s^2 = \sum_{i=1}^{k} (\bar{z}_i - \bar{z})^2 / (k - 1) - \hat{\sigma}^2 / k,$$

an adjustment of $\hat{\sigma}^2 / k$ from (7.21).

The smoothing parameter λ has to be chosen. Mathematical procedures exist, which enable an automatic choice to be made. For example, a so-called *pseudo-maximum likelihood* procedure (Habbema et al., 1974) was used to determine the value of λ (0.09) used in Figure 7.4. A value

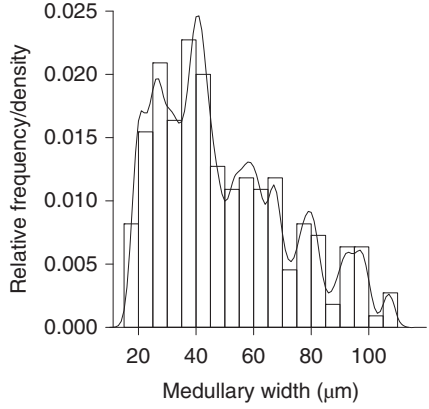

Figure 7.4 Medullary widths, in microns, of cat hairs (Source: Based on Peabody et al., 1983) and associated kernel density estimate with smoothing parameter equal to 0.09.

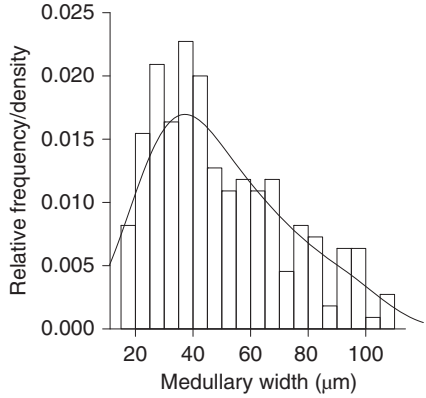

Figure 7.5 Medullary widths, in microns, of cat hairs (Source: Based on Peabody et al., 1983) and associated kernel density estimate with smoothing parameter equal to 0.50.

of λ equal to 0.50 was used to produce the curve in Figure 7.5 to illustrate the effect that a larger value of λ produces a smoother curve. Functions for kernel smoothing are available in R.

The choice of λ has to be made bearing in mind that the aim of the analysis is to provide a value V for the evidence in a particular case, as represented by the likelihood ratio. Using the kernel density estimation procedure an expression for V is derived; see (7.24). An investigation of the variation in V as λ varies is worthwhile. If V does not vary greatly as λ varies, then a precise value for λ is not necessary. For example, it is feasible to choose λ subjectively by comparing the density estimate curve $\hat{f}$ obtained for various values of λ with the histogram of the data. The value that provides the

best visual fit can then be chosen. Alternatively, from a scientist's personal experience of the distribution of the measurements on the characteristic of interest, it may be thought that certain possible values are not fully represented in the data set D available for estimation. In such a situation a larger value of λ may be chosen in order to provide a smoother curve, more representative of the scientist's experience. The subjective comparison of several plots of the data, produced by smoothing by different amounts, may well help to give a greater understanding of the data than the consideration of one curve, produced by an automatic method.

The choice of λ is also sensitive to outlying observations. The original cat hair data included one hair with a medullary width over 139 microns, the next largest being under 108 microns. The value of λ chosen by the automatic pseudo-maximum likelihood procedure was 0.35, a value that produced a very different estimate of the probability density function from that produced by the value of λ of 0.09 when the data set was modified as has been done by replacing the value of 139 microns by a value of 63 microns. The choice of λ is also difficult if the data are presented in grouped form as is the case with the glass data (Table 7.11). In this case, the value of λ was chosen subjectively, see Figures 7.6 and 7.7 with values of λ of 0.025 and 0.25.

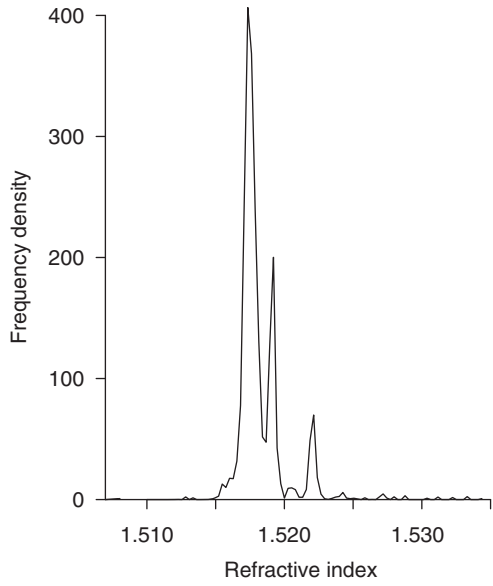

Figure 7.6 Kernel density estimate with smoothing parameter 0.025 of refractive index measurements from 2269 fragments of float glass from buildings (Source: Modified from Lambert and Evett, 1984).

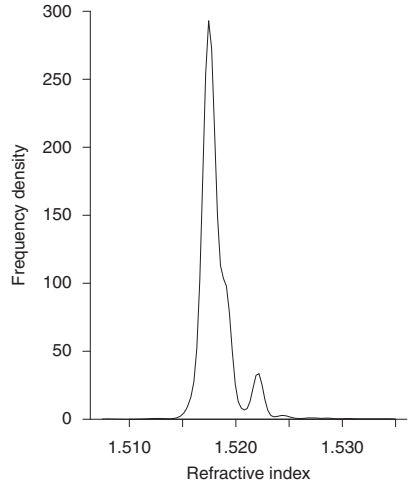

Figure 7.7 Kernel density estimate with smoothing parameter 0.25 of refractive index measurements from 2269 fragments of float glass from buildings (Source: Modified from Lambert and Evett, 1984).

7.5.2 Kernel Density Estimation for Between-Source Data

If the assumption of a Normal distribution for θ is thought unrealistic, the argument may be modified for a general distribution for θ using a kernel density estimation as described by Chan and Aitken (1989) for cat hairs, by Berry (1991) and Berry et al. (1992) for DNA profiling.

More recent examples are presented in Meuwly (2001), Gonzalez-Rodriguez et al. (2005), Neumann et al. (2006), Gonzalez-Rodriguez et al. (2007), Davis et al. (2012), and Ali et al. (2015).

An application to the evaluation of fibre evidence in which the marginal distribution of the recovered measurements y was estimated by a kernel density function was given by Evett et al. (1987) and a rather more elaborate treatment was given by Wakefield et al. (1991). This was a bivariate case involving colour measurements. Further details of these ideas are given in Section 7.6 where they are applied to multivariate data.

The method described here is applicable to situations in which the data are univariate and for which there are two components of variability, that within a particular source (e.g. window or cat) and that between different sources (e.g. windows or cats).

Consider the numerator in the original expression (7.8) for V, namely,

$$\int f(\mathbf{y} \mid \theta) f(\mathbf{x} \mid \theta) f(\theta) d\theta.$$

Given the value for θ, the distribution of $(\bar{X} - \bar{Y})$ is $N(0, a^2\sigma^2)$ and the distribution of W, given θ, is $N(\theta, \sigma^2/(m+n))$. If a change in the numerator is made from $(\bar{x}, \bar{y})$ to $(\bar{x} - \bar{y}, w)$ then V may be written as

$$\frac{\frac{1}{a\sigma}\exp\{-\frac{(\bar{x}-\bar{y})^2}{2a^2\sigma^2}\}\int \frac{(m+n)^{1/2}}{\sigma}\exp\{-\frac{(w-\theta)^2(m+n)}{2\sigma^2}\}f(\theta)d\theta}{\int \frac{\sqrt{m}}{\sigma}\exp\{-\frac{(\bar{x}-\theta)^2 m}{2\sigma^2}\}f(\theta)d\theta \int \frac{\sqrt{n}}{\sigma}\exp\{-\frac{(\bar{y}-\theta)^2 n}{2\sigma^2}\}f(\theta)d\theta} \tag{7.23}$$

(Lindley, 1977c). The probability density function, $f(\theta)$, for θ was previously assumed to be Normal. If this is thought to be unrealistic, the probability density function may be estimated by kernel density estimation.

The expression for V in (7.23) may be evaluated when $f(\theta)$ is replaced by the expression in (7.22). Some straightforward, but tedious, mathematics gives the result, a kernel approach to the calculation of a likelihood ratio, that

$$V = \frac{K\exp\{-\frac{(\bar{x}-\bar{y})^2}{2a^2\sigma^2}\}\sum_{i=1}^{k}\exp\{-\frac{(m+n)(w-z_i)^2}{2[\sigma^2+(m+n)s^2\lambda^2]}\}}{\sum_{i=1}^{k}\exp\{-\frac{m(\bar{x}-z_i)^2}{2(\sigma^2+ms^2\lambda^2)}\}\sum_{i=1}^{k}\exp\{-\frac{n(\bar{y}-z_i)^2}{2(\sigma^2+ns^2\lambda^2)}\}}, \tag{7.24}$$

where

$$K = \frac{k\sqrt{(m+n)}\sqrt{(\sigma^2+ms^2\lambda^2)}\sqrt{(\sigma^2+ns^2\lambda^2)}}{a\sigma\sqrt{(mn)}\sqrt{\{\sigma^2+(m+n)s^2\lambda^2\}}}.$$

There are four factors, explicitly dependent on the data, in the expression for V which contribute to its overall value:

(a) $\exp\{-(\bar{x}-\bar{y})^2/(2a^2\sigma^2)\}$;

(b) $\sum_{i=1}^{k} \exp\{-(m+n)(w-z_i)^2/$
$2[\sigma^2+(m+n)s^2\lambda^2]\}$;

(c) $\sum_{i=1}^{k} \exp\{-m(\bar{x}-z_i)^2/2(\sigma^2+ms^2\lambda^2)\}$;

(d) $\sum_{i=1}^{k} \exp\{-n(\bar{y}-z_i)^2/2(\sigma^2+ns^2\lambda^2)\}$.

The first factor, (a), accounts for the difference between the control and recovered evidence. A large difference leads to a smaller value of V, a small difference to a larger value of V.

The second factor, (b), accounts for the location of the combined evidence in the overall distribution from the relevant population. If it is far from the centre of this distribution then V will be smaller than if it were close. This provides a measure of the rarity of the combined evidence.

The third and fourth factors, (c) and (d), account for the rarity or otherwise of the source and receptor evidence, separately. The further these are from the centre of the overall distribution the smaller the corresponding factor and the larger the value of V.

Notice, also, the difference between σ^2, which measures the variance within a particular source (e.g. window or cat), and s^2, which estimates the overall between-group variance.

7.5.3 Examples

7.5.3.1 Medullary Widths of Cat Hairs

Consider a crime in which a cat is involved. For example, in a domestic burglary, there may have

been a cat at the crime scene. A PoI is identified who has cat hairs on their clothing. A full assessment of the evidence would require consideration of the PoI's explanation for the presence of these hairs and of the probabilities of transfer of cat hairs from the scene of the crime and from elsewhere. Such issues are not debated here. Measurements are made of the medullary widths, among other characteristics, of these hairs and of a sample of hairs from the domestic cat. Let $\bar{x}$ denote the mean of m hairs from the source (the domestic cat), let $\bar{y}$ denote the mean of n hairs from the receptor (the PoI's clothing). Some sample results for the value, V, as given by (7.24) of the evidence are given in Table 7.10 for various values of $\bar{x}$, $\bar{y}$, and

Table 7.10 Value of the evidence for the medullary width of cat hairs for various values of $\bar{x}$ and $\bar{y}$, the smoothing parameter λ and the within cat standard deviation σ; $m = n = 10$ throughout; $s = 23$ microns.

$\bar{x}$	$\bar{y}$	σ	V	
			$\lambda = 0.09$	$\lambda = 0.50$
15	15	10	16.50	12.01
15	25	10	1.39	0.782
15	35	10	9.81×10^{-4}	4.47×10^{-4}
110	110	10	84.48	53.61
50	50	10	6.97	6.25
50	50	16	3.86	3.93
50	50	5	16.14	12.48
50	55	10	3.75	3.54

σ. Variation in σ is given to illustrate the effect of changes in the variation within cats of medullary width on the value of the evidence. The value of the smoothing parameter λ has been taken to be 0.09 and 0.50 to illustrate variation in V with λ. The corresponding density estimate curves are shown in Figures 7.4 and 7.5, which can be used to assess the relative typicality of the evidence.

7.5.3.2 Refractive Index

These data, from Lambert and Evett (1984), are shown in Table 7.11 and illustrated in Figure 7.1. There are many coincident points and an automatic choice of λ is difficult and perhaps not desirable. Figures 7.6 and 7.7 show the kernel density estimate curves for λ equal to 0.025 and 0.25. The coincidence probabilities, from (3.8), and values V of the evidence, using (7.20) and (7.24), are given in Table 7.12.

Notice that, in general, the kernel approach leads to considerably higher values for V than does the Lindley approach. This arises from the more dispersed nature of the Lindley expression. Two examples in Table 7.12 show the failure of the coincidence probability approach. These are examples in which the separation of the control ($\bar{x}$) and recovered ($\bar{y}$) fragments is such that an approach based on coincidence probabilities would declare these two sets of fragments to have come from different windows. However, both the kernel and Lindley approaches give support to the proposition that they come from the same window.

Table 7.11 Refractive index of 2 269 fragments of float glass from buildings.

r.i.	Count	*r.i.*	Count	*r.i.*	Count	*r.i.*	Count
1.5081	1	1.5170	65	1.5197	7	1.5230	1
1.5119	1	1.5171	93	1.5198	1	1.5233	1
1.5124	1	1.5172	142	1.5199	2	1.5234	1
1.5128	1	1.5173	145	1.5201	4	1.5237	1
1.5134	1	1.5174	167	1.5202	2	1.5240	1
1.5143	1	1.5175	173	1.5203	4	1.5241	1
1.5146	1	1.5176	128	1.5204	2	1.5242	1
1.5149	1	1.5177	127	1.5205	3	1.5243	3
1.5151	1	1.5178	111	1.5206	5	1.5244	1
1.5152	1	1.5179	81	1.5207	2	1.5246	2
1.5153	1	1.5180	70	1.5208	3	1.5247	2
1.5154	3	1.5181	55	1.5209	2	1.5249	1
1.5155	5	1.5182	40	1.5211	1	1.5250	1
1.5156	2	1.5183	28	1.5212	1	1.5254	1
1.5157	1	1.5184	18	1.5213	1	1.5259	1
1.5158	7	1.5185	15	1.5215	1	1.5265	1
1.5159	13	1.5186	11	1.5216	3	1.5269	1
1.5160	6	1.5187	19	1.5217	4	1.5272	2
1.5161	6	1.5188	33	1.5218	12	1.5274	1
1.5162	7	1.5189	47	1.5219	21	1.5280	1
1.5163	6	1.5190	51	1.5220	30	1.5287	2
1.5164	8	1.5191	64	1.5221	25	1.5288	1
1.5165	9	1.5192	72	1.5222	28	1.5303	2
1.5166	16	1.5193	56	1.5223	13	1.5312	1
1.5167	15	1.5194	30	1.5224	6	1.5322	1
1.5168	25	1.5195	11	1.5225	3	1.5333	1
1.5169	49	1.5196	3	1.5226	5	1.5343	1

Source: Lambert and Evett (1984). Reproduced with permission of Elsevier.

Table 7.12 Coincidence probability and value of the evidence for the refractive index of glass (kernel and Lindley approaches) for various values of $\bar{x}$ and $\bar{y}$ and the smoothing parameter λ (for the kernel approach); $m = 10$, $n = 5$; within window standard deviation $\sigma = 0.000\,04$, between window standard deviation $\tau = 0.004$; overall mean $\mu = 1.5182$.

$\bar{x}$	$\bar{y}$	Coincidence probability	λ for kernel approach			Lindley
			0.025	0.05	0.25	
1.515 00	1.515 01	2.845×10^{-9}	17 889	7 055	2 810	226
1.516 00	1.516 01	2.643×10^{-3}	563	489	419	191
1.517 00	1.517 01	2.863×10^{-2}	54.3	52.4	48.9	172
1.518 00	1.518 01	3.083×10^{-2}	53.3	54.4	49.2	164
1.519 00	1.519 01	2.246×10^{-2}	70.0	69.2	102.4	167
1.520 00	1.520 01	8.536×10^{-9}	5 524	2 297	471.2	182
1.521 00	1.521 01	4.268×10^{-9}	13 083	4 381	1 397	210
1.522 00	1.522 01	1.321×10^{-2}	128	143	304	259
1.515 00	1.515 05	—	740	519	217	18.4
1.516 00	1.516 05	—	48.4	42.4	32.6	15.6
1.516 00	1.516 10	—	1.76×10^{-2}	1.74×10^{-2}	1.22×10^{-2}	6.30×10^{-3}
1.517 00	1.517 10	—	1.35×10^{-3}	1.42×10^{-3}	1.51×10^{-3}	5.69×10^{-3}

7.6 MULTIVARIATE ANALYSIS

7.6.1 Introduction

Multivariate data are becoming more prevalent in forensic science. Often more than one characteristic, or variable, is recorded for a piece of evidence. Glass fragments that are searched and recovered at a crime scene or drug samples that are seized under suspicion of containing illicit substances may be analysed and compared on the basis of a profile of the elemental compositions of chemical compounds as well as physical characteristics. A comprehensive review of statistical analysis for the evaluation of multivariate physicochemical data can be found in Zadora et al. (2014). Multivariate data may arise in other domains of forensic science, such as handwriting examination, where each handwritten character can be described by several variables (such as height, surface, or by Fourier descriptors as proposed by Marquis et al. (2005)).

One of the criticisms that can be addressed against the use of multivariate methods in forensic science is the lack of background information from which to estimate the model parameters. Forensic laboratories often need to deal with databases characterised by a complex dependence structure, with a large number of variables and multiple sources of variation. Score-based models are an attempt to reduce dimensionality, as they allow a reduction in the multivariate structure

to a univariate distance or similarity score (as is illustrated in Section 7.8). In other situations, provided variables are (at least roughly) independent, the likelihood ratio can be simplified to a product of univariate likelihood ratios. However, the hypothesis of independence is seldom guaranteed, these various characteristics may not be independent and it might be necessary to allow for the dependence among variables in the evaluation of evidence. An example of the importance of allowing for dependence between characteristics has been given in Section 3.6.3 where the characteristics are the refractive index and density of glass. The product of two separate significance probabilities was 0.0016, which may be thought to be highly significant. It was shown in Section 3.6.3 that when the dependence between two variables was included in the analysis, the significance probability was only 0.1225. The two characteristics were significant individually at the 5% level but together were not significant at the 10% level. It is possible to transform multivariate data, for example, using a method known as principal component analysis (Jolliffe, 1986). Such an approach in the context of footprints is discussed in Kennedy et al. (2003).

Some methods for the evaluation of the evidence in the form of multivariate data will be presented in this chapter. The likelihood ratio generalises the likelihood ratio in (7.17) and assumes that data are Normally distributed. As in the development of (7.17), there will be assumed multiple sources of

variation. In Section 7.6.2 a two-level model that accounts for within-source and between-source variation will be considered. A three level model that accounts also for variation due to measurement error will be considered in Section 7.6.5.

The ideas will be illustrated with an example concerning elemental concentrations of glass. The arithmetic involved is considerable. In order to have data that can be presented easily an example will be considered in Section 7.6.4 in which only three variables are considered and in which there are only two control and two recovered items. Two sets of results will be calculated, one for when the control and recovered items come from the same source and one for when the control and recovered items come from different sources.

Other methods relying on frequentist ideas might be considered. A significance test based on Hotelling's T^2-statistic has been proposed by Curran et al. (1997a,b).

7.6.2 Multivariate Two-Level Models

Let Ω denote a population of p characteristics of items of a particular evidential type. Continuous measurements of these characteristics are available on a random sample of m members from Ω with $n(\geq 2)$ replicate measurements on each of the m members. In the case of glass data, items are fragments and the characteristics measured are concentrations in elemental composition. The

background data are denoted $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$, $i = 1, \ldots, m$ and $j = 1, \ldots, n$ with $\bar{\mathbf{x}}_i = \sum_{i=1}^{n} \mathbf{x}_{ij}$.

As in Section 7.3.2 the model discussed here assumes two sources of variation, that between replicates within the same group or source (known as within-group variation) and that between groups or sources (known as between-group variation). It is assumed that both the variation within-group and the variation between-group are Normally distributed.

Within-group: Denote the mean vector within group i by $\boldsymbol{\theta}_i$ and the within-group covariance matrix by U. The subscript i is omitted from the covariance matrix to indicate that it is assumed that the within-group variability is constant over all groups. This is an extension of the assumption made in standard univariate analysis of variance techniques. This assumption will be relaxed in Section 7.6.2.4. Then, given $\boldsymbol{\theta}_i$ and U, the distribution of $\mathbf{X}_{ij}$ (a $p \times 1$ column vector indicating the p characteristics of the j-th member of the i-th group) is taken to be Normal, where the notation is the same as in Section A.3.9:

$$(\mathbf{X}_{ij} \mid \boldsymbol{\theta}_i, U) \sim N(\boldsymbol{\theta}_i, U),\ i = 1, \ldots, m;\ j = 1, \ldots, n.$$

Between-group: Denote the mean vector between groups by $\boldsymbol{\mu}$ and the between-group covariance matrix by C. The distribution of the $\boldsymbol{\theta}_i$, as measures of between-source variability, is taken to be Normal:

$$(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, C) \sim N(\boldsymbol{\mu}, C).$$

This distribution arises as the model is assumed to be a so-called *random effects* model. The different groups in the population database are thought of as a random sample from a larger population (or super-population). Thus, measurements on the groups in the (population) database are a random sample from a larger population and thus have variability. The term 'multivariate random effects' arises because the data are multivariate.

There are multivariate data from a crime scene, which are assumed to come from one source, and multivariate data from a PoI, which are assumed to come from one source. These sources may or may not be the same source. A likelihood ratio is derived to measure the support for the proposition that the sources are the same as opposed to the proposition that the sources are different.

Control data are denoted $\mathbf{y}_1$ and recovered data are denoted $\mathbf{y}_2$. Note that, contrary to the convention elsewhere in the book that $\mathbf{x}$ denotes control and $\mathbf{y}$ denotes recovered data, the control and recovered data are distinguished by subscripts, and $\mathbf{y}$ denotes control or recovered data. Also $\mathbf{x}$ refers to training data whereas elsewhere (e.g. Section 7.5.1) $\mathbf{z}$ refers to training data. This is for ease of notation in the description of the models and in comparison with Aitken and Lucy (2004).

Let there be n_1 observations at the crime scene (recovered data), with vectors of measurements $\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}$ and n_2 observations from a PoI (control data), with vectors of measurements $\mathbf{y}_{21}, \ldots, \mathbf{y}_{2n_2}$, where n_1 is not necessarily equal

to n_2. The means of these two sets of observations are

$$\bar{\mathbf{y}}_1 = \sum_{j=1}^{n_1} \mathbf{y}_{1j}/n_1, \text{ and } \bar{\mathbf{y}}_2 = \sum_{j=1}^{n_2} \mathbf{y}_{2j}/n_2.$$

The distributions of the means $(\bar{\mathbf{Y}}_l; l = 1, 2)$ of the measurements on the recovered and control data, conditional on the source, are also taken to be Normal, with means $\boldsymbol{\theta}_l$ and covariance matrix D_l where $D_1 = n_1^{-1}U$ and $D_2 = n_2^{-1}U$. Thus

$$(\bar{\mathbf{Y}}_l \mid \boldsymbol{\theta}_l, D_l) \sim N(\boldsymbol{\theta}_l, D_l); \;\; l = 1, 2.$$

Then it can be shown that

$$(\bar{\mathbf{Y}}_l \mid \boldsymbol{\mu}, C, D_l) \sim N(\boldsymbol{\mu}, C + D_l); \;\; l = 1, 2.$$

This is the multivariate generalisation of (7.15).

7.6.2.1 Parameter Estimation

Information on the overall vector mean $\boldsymbol{\mu}$, the within- (U), and between-source (C) covariance matrices can be assessed on the basis of background information obtained from a suitable database $\mathbf{x} = \{x_{ij}, i = 1, \ldots, m, j = 1, \ldots, n\}$ where m is the number of groups and n is the number of members of each group, assumed constant between groups, with $mn = N$. The mean $\boldsymbol{\mu}$ is estimated by $\bar{\mathbf{x}}$, the mean vector over all groups, that is,

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{x}_{ij}. \qquad (7.25)$$

The within-group covariance matrix U is estimated from the background data $\{\mathbf{x}\}$ by

$$\widehat{U} = \frac{S_w}{(N-m)} \tag{7.26}$$

where

$$S_w = \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T.$$

The between-group covariance matrix C is estimated from the background data $\{\mathbf{x}_{ij}\}$ by

$$\widehat{C} = \frac{S^*}{m-1} - \frac{S_w}{n(N-m)}, \tag{7.27}$$

where

$$S^* = \sum_{i=1}^{m} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T.$$

7.6.2.2 Likelihood Ratio Using a Multivariate Random Effects Model and Assumptions of Normality

The value of the evidence $(\mathbf{y}_l; l = 1, 2)$ is then the ratio of two probability density functions, evaluated at the point $(\mathbf{y}_1, \mathbf{y}_2)$ given two competing propositions H_p and H_d. For the calculations in the numerator (i.e. H_p holds) it is assumed that the control and recovered items come from the same source and the means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are equal. For the calculations in the denominator (i.e. H_d holds), it is assumed that the control and recovered items

come from different sources and the means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are not equal.

First, consider the numerator where $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}$, say, which is unknown. The parameter $\boldsymbol{\theta}$ can be eliminated by integration, analogous to the approach of Section 7.3, to obtain a probability density function $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, U, C)$ given by

$$\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, D_1) f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}.$$

The component probability density functions are multivariate Normal. The expressions for $f(\mathbf{y}_1 \mid \boldsymbol{\theta}, D_1), f(\mathbf{y}_2 \mid \boldsymbol{\theta}, D_2)$, and $f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C)$ earlier are obtained by appropriate substitutions in the general formula (A.27).

The integral can then be shown to be equal to

$$\begin{aligned} f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, U, C) = & \mid 2\pi U \mid^{-\frac{1}{2}(n_1+n_2)} \mid 2\pi C \mid^{-1/2} \\ & \times \mid 2\pi \{(n_1+n_2)U^{-1} + C^{-1}\}^{-1} \mid^{1/2} \\ & \exp\left\{-\frac{1}{2}(H_1 + H_2 + H_3)\right\} \end{aligned} \tag{7.28}$$

where

$$H_1 = \sum_{l=1}^{2} \operatorname{tr}(S_l U^{-1}), \tag{7.29}$$

$$H_2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^T \left(\frac{U}{n_1+n_2} + C\right)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}), \tag{7.30}$$

$$H_3 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \tag{7.31}$$

$$\bar{\mathbf{y}} = (n_1\bar{\mathbf{y}}_1 + n_2\bar{\mathbf{y}}_2)/(n_1 + n_2),$$

$$S_l = \sum_{j=1}^{n_l}(\mathbf{y}_{lj} - \bar{\mathbf{y}}_l)(\mathbf{y}_{lj} - \bar{\mathbf{y}}_l)^T,$$

(Aitken and Lucy, 2004). Note that the term tr in (7.29) denotes the trace of a matrix, Section B.1.1. The exponential term in (7.28) is a combination of three terms, H_3 that accounts for the difference $(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$ between the means of the measurements on the control and recovered items, H_2 that accounts for their rarity (as measured by the distance of the mean weighted by sample sizes from $\boldsymbol{\mu}$), and H_1 that accounts for internal variability.

Second, consider the denominator where $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. The probability density function $f_1(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, U, C)$ is given by

$$\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}$$

$$\times \int f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta},$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are taken to be independent as the data are assumed to be from different sources. The integral

$$\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}$$

can be shown to be equal to

$$f(\mathbf{y}_1 \mid \boldsymbol{\mu}, U, C)$$
$$= \mid 2\pi U \mid^{-n_1/2} \mid 2\pi C \mid^{-1/2} \mid 2\pi(n_1 U^{-1} + C^{-1})^{-1} \mid^{1/2}$$

$$\times \exp\left\{-\frac{1}{2}\text{tr}(S_1 U^{-1}) - \frac{1}{2}(\bar{\mathbf{y}}_1 - \boldsymbol{\mu})^T \left(\frac{U}{n_1} + C\right)^{-1} (\bar{\mathbf{y}}_1 - \boldsymbol{\mu})\right\}, \tag{7.32}$$

with an analogous result for

$$\int f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}.$$

The value of the evidence is the ratio of $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, U, C)$ to the product of $\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}$ and $\int f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}$. After some manipulations it can be shown this is equal to the ratio of

$$\mid 2\pi\{(n_1 + n_2)U^{-1} + C^{-1}\}^{-1} \mid^{1/2} \exp\left\{-\frac{1}{2}(H_2 + H_3)\right\} \tag{7.33}$$

to

$$\begin{aligned} &\mid 2\pi C \mid^{-1/2} \mid 2\pi(n_1 U^{-1} + C^{-1})^{-1} \mid^{1/2} \\ &\quad \mid 2\pi(n_2 U^{-1} + C^{-1})^{-1} \mid^{1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}(H_4 + H_5)\right\} \end{aligned} \tag{7.34}$$

where

$$\begin{aligned} H_4 = (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^T &\{(D_1 + C)^{-1} + (D_2 + C)^{-1}\} \\ &(\boldsymbol{\mu} - \boldsymbol{\mu}^*), \end{aligned} \tag{7.35}$$

$$H_5 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2 + 2C)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \tag{7.36}$$

$$\boldsymbol{\mu}^* = \{(D_1 + C)^{-1} + (D_2 + C)^{-1}\}^{-1} \{(D_1 + C)^{-1} \bar{\mathbf{y}}_1 + (D_2 + C)^{-1} \bar{\mathbf{y}}_2\},$$

$D_1 = U/n_1$ and $D_2 = U/n_2$.

Alternatively, a maximum likelihood approach has been proposed by Ommen and Saunders (2018).

7.6.2.3 Likelihood Ratio Using a Multivariate Random Effects Model and Non-normal Between-Source Variation

A multivariate normal distribution for the vector mean $\boldsymbol{\theta}$ may not always be a reasonable assumption. The assumption of normality can be removed by considering a kernel density estimate for the between-group distribution (as in Section 7.5.2). Given a data set $D = \{\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m\}$, which in this case will be taken to be the vector of group means $\bar{\mathbf{x}}_i$, $i = 1, \ldots, m$, the kernel density function is taken to be a multivariate Normal density function, with a mean at $\bar{\mathbf{x}}_i$ and covariance matrix $h^2 C$, and denoted by $K(\boldsymbol{\theta} \mid \bar{\mathbf{x}}_i, C, h)$ where

$$K(\boldsymbol{\theta} \mid \bar{\mathbf{x}}_i, C, h) = \frac{(2\pi)^{-p/2} \mid C \mid^{-1/2}}{h^p} \exp\left\{-\frac{1}{2} h^{-2} (\boldsymbol{\theta} - \bar{\mathbf{x}}_i)^T C^{-1} (\boldsymbol{\theta} - \bar{\mathbf{x}}_i)\right\}.$$

The estimate $f(\boldsymbol{\theta} \mid \bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m, C, h)$ of the overall probability density function is then

$$f(\boldsymbol{\theta} \mid \bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m, C, h) = \frac{1}{m} \sum_{i=1}^{m} K(\boldsymbol{\theta} \mid \bar{\mathbf{x}}_i, C, h), \tag{7.37}$$

where the smoothing parameter h can be estimated as

$$\hat{h} = \left(\frac{4}{2p+1} \right)^{1/(p+4)} m^{-1/(p+4)} \tag{7.38}$$

(Silverman, 1986; Scott, 1992).

The numerator of the likelihood ratio, for which hypothesis H_p is assumed true, can be shown to be given by (see (Aitken and Lucy, 2004))

$$\begin{aligned} f_0(\mathbf{y}_1, \mathbf{y}_2 \mid U, C) &= (2\pi)^{-p} \mid D_1 \mid^{-\frac{1}{2}} \\ &\quad \mid D_2 \mid^{-\frac{1}{2}} \mid \\ &\quad \mid C \mid^{-\frac{1}{2}} (mh^p)^{-1} \mid D_1^{-1} + D_2^{-1} + (h^2 C)^{-1} \mid^{-\frac{1}{2}} \\ &\quad \times \exp\left\{ -\frac{1}{2} H_3 \right\} \times \sum_{i=1}^{m} \exp\left\{ -\frac{1}{2} H_i \right\}, \end{aligned} \tag{7.39}$$

where H_3 is as in (7.31),

$$\begin{aligned} H_i = (\mathbf{y}^* - \bar{\mathbf{x}}_i)^T \{ (D_1^{-1} + D_2^{-1})^{-1} + (h^2 C) \}^{-1} \\ (\mathbf{y}^* - \bar{\mathbf{x}}_i), \quad i = 1, \ldots, m, \end{aligned}$$

and $\mathbf{y}^* = (D_1^{-1} + D_2^{-1})^{-1} (D_1^{-1} \bar{\mathbf{y}}_1 + D_2^{-1} \bar{\mathbf{y}}_2)$.

The denominator of the likelihood ratio, for which hypothesis H_d is assumed true, can be

shown to be given by

$$f_1(\mathbf{y}_1, \mathbf{y}_2 \mid U, C) = (2\pi)^{-p} \mid C \mid^{-1} (mh^p)^{-2} \prod_{l=1}^{2} \left[\mid D_l \mid^{-\frac{1}{2}} \mid D_l^{-1} + (h^2 C)^{-1} \mid^{-\frac{1}{2}} \sum_{i=1}^{m} \exp\left\{ -\frac{1}{2} H_{li} \right\} \right], \tag{7.40}$$

with

$$H_{li} = (\bar{\mathbf{y}}_l - \bar{\mathbf{x}}_i)^T (D_l + h^2 C)^{-1} (\bar{\mathbf{y}}_l - \bar{\mathbf{x}}_i), \quad l = 1, 2 ; \quad i = 1, \ldots, m.$$

The likelihood ratio is then the ratio of (7.39)–(7.40) and is

$$\frac{\mid C \mid^{\frac{1}{2}} mh^p \mid D_1^{-1} + D_2^{-1} + (h^2 C)^{-1} \mid^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} H_3 \right\} \sum_{i=1}^{m} \exp\left\{ -\frac{1}{2} H_i \right\}}{\prod_{l=1}^{2} \left[\mid D_l^{-1} + (h^2 C)^{-1} \mid^{-\frac{1}{2}} \sum_{i=1}^{m} \exp\left\{ -\frac{1}{2} H_{li} \right\} \right]} \tag{7.41}$$

A kernel density estimation procedure with an unconstrained bandwidth matrix is recommended by Neocleous et al. (2011) for the evaluation of evidence in the form of compositional data.[1]

[1] Compositional data are multivariate data whose individual components add up to a fixed, known, constant, usually 1 or 100%. An example is the elemental composition of glass where the proportions of the elements which form the composition of the glass add up to 100%.

Alternatively, Pedroso et al. (2016) proposed the use of a Gaussian mixture model for the between-source variation, where observations are assumed to be generated from a mixture of a finite number of Normal densities with unknown parameters estimated by a maximum likelihood approach.

7.6.2.4 Non-constant Within-Group Covariance Matrix

There may be domains where a constant variability within groups is difficult to justify. One example is given by handwriting comparison, where it can be observed that each writer is characterised by a particular variability (Marquis et al., 2006). To account for a non-constant variability within sources, Bozza et al. (2008) proposed a two-level model, where the distribution of the $\mathbf{X}_{ij}$ is taken to be Normal, with mean given by the mean vector within group $\boldsymbol{\theta}_i$, but with a non-constant within-group covariance matrix U_i,

$$(\mathbf{X}_{ij} \mid \boldsymbol{\theta}_i, U_i) \sim N(\boldsymbol{\theta}_i, U_i), \;\; i = 1, \ldots, m; \;\; j = 1, \ldots, n.$$

The distribution of the $\boldsymbol{\theta}_i$ is taken as in Section 7.6.2, whilst an inverse Wishart distribution (Section A.3.11) is taken for the within-group covariance matrix U_i,

$$(U_i \mid \Sigma, \nu) \sim W^{-1}(\Sigma, \nu),$$

where the scale matrix Σ is taken so that the prior mean of the $\{U_i\}$ is taken to be equal to

the common within-group covariance matrix estimated from the background data and ν is the prior degrees of freedom. A two-level multivariate random effects model with a non-constant within-group covariance matrix has been adopted also by Ommen et al. (2017).

Consider the numerator where $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}$, say, and $U_1 = U_2 = U$, which are unknown. The probability density function $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, C, \Sigma, \nu)$ under H_p can be obtained as

$$\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta}, U \mid \boldsymbol{\mu}, C, \Sigma, \nu) d(\boldsymbol{\theta}, U). \tag{7.42}$$

Consider the denominator, where $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ and $U_1 \neq U_2$. The probability density function $f_1(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, C, \Sigma, \nu)$ under H_d is given by

$$\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta}, W \mid \boldsymbol{\mu}, C, \Sigma, \nu) d(\boldsymbol{\theta}, U) \times \int f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta}, W \mid \boldsymbol{\mu}, C, \Sigma, \nu) d(\boldsymbol{\theta}, U). \tag{7.43}$$

However, the integrations in (7.42) and (7.43) do not have an analytical solution. The marginal likelihoods $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, C, \Sigma, \nu)$ and $f_1(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\mu}, C, \Sigma, \nu)$ can be obtained by means of Markov chain Monte Carlo methods, as in Bozza et al. (2008) where a Gibbs sampling algorithm was applied to the set of the conditional densities. A summary of fundamental elements of Bayesian computation and Markov chain simulation can be found in Gelman et al. (2014).

7.6.3 A Note on Sensitivity

The described procedures might be sensitive to changes in the control and recovered measurements. A sensitivity analysis may be conducted in terms of the percentages of false negatives and false positives that are obtained whenever measurements originating from a common or from different sources are compared (e.g. Aitken and Lucy (2004), Bozza et al. (2008)). The scientist may in this way focus on how many times a particular likelihood ratio – obtained for a particular setting (e.g. same or different source) – points in the wrong direction (i.e. supports the first proposition instead of the second, or *vice versa*). This is not to provide a distribution for the likelihood ratio. Given the available measurements, background information, statistical model, and prior knowledge, the best estimate of the value of the evidence is given by a single number. The purpose of considering false positives and false negatives is to inform about the potential of misleading evidence (Taroni et al., 2016) (see Chapter 8). In fact, as reported by Taylor et al. (2016c),

> *The key point is that it is not our belief (i.e. our probability) that is more robust, but the basis of this belief (i.e. the amount of knowledge, the data). (p. 408)*

This use of the term 'sensitivity' should be contrasted with the use given in Section 2.2.2 for the performance of a test.

7.6.4 Case Study for Two-Level Data

The example that is used to illustrate the methods in Sections 7.6.2.2 and 7.6.2.3 is based on a database consisting of the measurements of the elemental concentration on glass fragments from several ($m = 62$) windows (Aitken and Lucy, 2004). Three derived variables were retained because they were considered the most discriminatory (i.e. the ratios *Ca*/*K*, *Ca*/*Si*, *Ca*/*Fe*). A logarithmic transformation was made to reduce positive skewness and to make the normality assumption more valid.

The overall mean elemental concentration $\boldsymbol{\mu}$ is estimated by the overall sample mean $\bar{\mathbf{x}}$ for the three log elemental ratios as in (7.25)

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = (4.20, -0.75, 2.77)^T.$$

The within-group covariance matrix U is estimated by $\widehat{U}$ as in (7.26)

$$\widehat{U} = \begin{pmatrix} 1.68 \times 10^{-2} & 2.66 \times 10^{-5} & 2.21 \times 10^{-4} \\ 2.66 \times 10^{-5} & 6.53 \times 10^{-5} & 7.40 \times 10^{-6} \\ 2.21 \times 10^{-4} & 7.40 \times 10^{-6} & 1.33 \times 10^{-3} \end{pmatrix}.$$

The between-group covariance matrix C is estimated by $\widehat{C}$ as in (7.27)

$$\widehat{C} = \begin{pmatrix} 7.06 \times 10^{-1} & 9.88 \times 10^{-2} & -4.63 \times 10^{-2} \\ 9.88 \times 10^{-2} & 6.21 \times 10^{-2} & 6.96 \times 10^{-3} \\ -4.63 \times 10^{-2} & 6.96 \times 10^{-3} & 1.01 \times 10^{-1} \end{pmatrix}.$$

The measurements $\mathbf{y}_{11}, \mathbf{y}_{12}$ on the two control fragments are taken to be the same throughout. These are

$$\mathbf{y}_{11} = \begin{pmatrix} 3.774 \\ -0.891 \\ 2.620 \end{pmatrix}, \ \mathbf{y}_{12} = \begin{pmatrix} 3.939 \\ -0.893 \\ 2.639 \end{pmatrix}.$$

The mean vector of these two vectors is obtained by taking the mean of the three components. These three means together give the mean vector. This may be denoted $\bar{\mathbf{y}}_1$ and is

$$\bar{\mathbf{y}}_1 = (3.856, -0.892, 2.629)^T.$$

There will be two sets of recovered fragments. One will be used for the evaluation of the evidence when the control and recovered fragments come from the same source, and one will be used for the evaluation of the evidence when the control and recovered fragments come from different sources.

The measurements $\mathbf{y}_{21}, \mathbf{y}_{22}$ on the two recovered fragments taken to be from the same source as the control fragments are

$$\mathbf{y}_{21} = \begin{pmatrix} 3.844 \\ -0.910 \\ 2.654 \end{pmatrix}, \ \mathbf{y}_{22} = \begin{pmatrix} 3.725 \\ -0.898 \\ 2.619 \end{pmatrix},$$

with mean

$$\bar{\mathbf{y}}_2 = (3.784, -0.9041, 2.6368)^T.$$

The measurements $\mathbf{y}_{31}, \mathbf{y}_{32}$ on the two recovered fragments taken to be from a different source to the control fragments are

$$\mathbf{y}_{31} = \begin{pmatrix} 4.077 \\ -0.835 \\ 2.739 \end{pmatrix}, \ \mathbf{y}_{32} = \begin{pmatrix} 4.109 \\ -0.819 \\ 2.796 \end{pmatrix},$$

with mean

$$\bar{\mathbf{y}}_3 = (4.0933, -0.8268, 2.7674)^T.$$

These are denoted with a subscript 3 to distinguish them from those recovered fragments, with a subscript 2, deemed to be from the same source as the control fragments, with a subscript 1.

7.6.4.1 Normal Density for Between-Group Distribution

First, consider the case where the control and recovered fragments are chosen to come from the same source.

Consider the determination of the numerator. The first term in (7.33) involves inverses of covariance matrices and a square root of a determinant of a function of the two covariance matrices U and C and gives

$$| \, 2\pi\{(n_1 + n_2)U^{-1} + C^{-1}\}^{-1} \, |^{1/2} = 7.467 \times 10^{-5}.$$

The term H_2 (7.30) measures the difference between the overall mean of the control and recovered fragments and the overall population mean. In this example,

$$H_2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^T \left(\frac{U}{n_1 + n_2} + C \right)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}) = 0.6734.$$

The term H_3 (7.31) measures the difference between the means of the control and recovered fragments. Note that, for the example under discussion, $n_1 = n_2 = 2$ and so $D_1 + D_2 = U$. In this example,

$$H_3 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = 2.5716.$$

So, $\exp\left\{ -\frac{1}{2}(H_2 + H_3) \right\} = 0.1974.$

All these terms can be put together to give the numerator of the likelihood ratio, (7.33), which is

$$(7.467 \times 10^{-5}) \times 0.1974 = 1.474 \times 10^{-5}.$$

Now consider the denominator. The various terms in (7.34) which involve determinants, inverses of covariance matrices and determinants of functions of the two covariance matrices U and C take the following values:

$$\begin{aligned} &| 2\pi C |^{-1/2} = 1.0993, \\ &| 2\pi (n_1 U^{-1} + C^{-1})^{-1} |^{1/2} \\ &\quad = | 2\pi (n_2 U^{-1} + C^{-1})^{-1} |^{1/2} \\ &\quad = 0.000\,21. \end{aligned}$$

(Note that in this example $n_1 = n_2 = 2$.) From (7.35)

$$H_4 = (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^T\{(D_1 + C)^{-1} + (D_2 + C)^{-1}\}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)$$
$$= 1.3437.$$

From (7.36)

$$H_5 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T(D_1 + D_2 + 2C)^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$
$$= 0.0037.$$

So, $\exp\left\{-\frac{1}{2}(H_4 + H_5)\right\} = 0.5098$.

All these terms can be put together to give the denominator of the likelihood ratio, which is

$$1.0993 \times 0.000\,21 \times 0.000\,21 \times 0.5098$$
$$= 2.471 \times 10^{-8}.$$

The value of the evidence is then

$$V = \frac{1.474 \times 10^{-5}}{2.471 \times 10^{-8}} = 596.47.$$

The likelihood ratio correctly supports the proposition according to which the control and recovered fragments originate from the same source.

Now, consider the case where the control and recovered fragments are chosen to come from different sources. Many of the terms and calculations will be the same as or similar to the calculations done for the example where the

control and recovered fragments were chosen to come from the same sources. This is because the control group has been chosen to be the same in both cases and because the covariance matrices U and C and the sample sizes are the same.

Consider the determination of the numerator. The first term in (7.33) which involves inverses of covariance matrices and a square root of a determinant of a functions of the two covariance matrices U and C is the same as before, namely,

$$| 2\pi\{(n_1 + n_2)U^{-1} + C^{-1}\}^{-1} |^{1/2} = 7.467 \times 10^{-5}.$$

The term H_2 (7.30), which measures the difference between the overall mean of the control and recovered fragments and the overall population mean, is now equal to

$$H_2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^T \left(\frac{U}{n_1 + n_2} + C \right)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}) = 0.2881.$$

The term H_3 (7.31), which measures the difference between the means of the control and recovered fragments, is now equal to

$$H_3 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = 80.007.$$

So, $\exp\left\{-\frac{1}{2}(H_2 + H_3)\right\} = 3.666 \times 10^{-18}$.

All these terms can be put together to give the numerator of the likelihood ratio, which is

$$(7.467 \times 10^{-5}) \times 3.666 \times 10^{-18} = 2.737 \times 10^{-22}.$$

Now consider the denominator. The various terms in (7.34) which involve determinants,

inverses of covariance matrices and determinants of functions of the two covariance matrices U and C are the same as before, namely,

$$\begin{aligned} &| 2\pi C|^{-1/2} = 1.0993, \\ &| 2\pi(n_1 U^{-1} + C^{-1})^{-1} |^{1/2} \\ &= | 2\pi(n_2 U^{-1} + C^{-1})^{-1} |^{1/2} \\ &= 0.000\,21. \end{aligned}$$

The term H_4 (7.35) is now equal to

$$\begin{aligned} H_4 &= (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^T \{(D_1 + C)^{-1} + (D_2 + C)^{-1}\}(\boldsymbol{\mu} - \boldsymbol{\mu}^*) \\ &= 0.5754. \end{aligned}$$

The term H_5 (7.36) is now equal to

$$\begin{aligned} H_5 &= (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2 + 2C)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) \\ &= 0.1697. \end{aligned}$$

So, $\exp\left\{-\frac{1}{2}(H_4 + H_5)\right\} = 0.6889$.

All these terms can be put together to give the denominator of the likelihood ratio, which is

$$\begin{aligned} &1.0993 \times 0.000\,21 \times 0.000\,21 \times 0.6889 \\ &= 3.34 \times 10^{-8}. \end{aligned}$$

The value of the evidence is then

$$V = \frac{2.733 \times 10^{-22}}{3.34 \times 10^{-8}} = 8.18 \times 10^{-15}.$$

The likelihood ratio correctly supports the proposition according to which the control and recovered fragments originate from different sources.

7.6.4.2 Kernel Density for Between-Group Distribution

First, consider the case where the control and recovered fragments are chosen to come from the same source.

Consider the determination of the numerator of the likelihood ratio where the control and recovered fragments are assumed to (i.e. hypothesis H_p is true). The first terms in the numerator that involve the square root of the determinant of the covariance matrix C and the square root of the inverse of the determinant of a function of the two covariance matrices U and C are

$$| C |^{1/2} = 0.0577,$$
$$| D_1^{-1} + D_2^{-1} + (h^2 C)^{-1} |^{-1/2} = 4.665 \times 10^{-6}.$$

The smoothing parameter h can be estimated as in (7.38) and equals 0.511 94. In this way

$$mh^p = 8.3189,$$

where $p = 3$ and $m = 62$.

The term H_3 has been calculated in Section 7.6.4.1, so that $\exp\left\{-\frac{1}{2}H_3\right\}$ equals 0.276. The expression

$$\sum_{i=1}^{m} \exp\left\{-\frac{1}{2}H_i\right\}$$

equals 7.237.

Consider the determination of the denominator of the likelihood ratio where the control and recovered fragments are assumed to originate from different sources (i.e. hypothesis H_d is true).

The first term in the denominator that involves inverses of covariance matrices and a square root of a determinant of a functions of the two covariance matrices U and C is the same as before, namely,

$$\begin{aligned} \mid D_1^{-1} + (h^2C)^{-1} \mid^{-1/2} &= \mid D_2^{-1} + (h^2C)^{-1} \mid^{-1/2} \\ &= 1.29 \times 10^{-5}. \end{aligned}$$

Note that in this example $n_1 = n_2$, so that $D_1 = D_2$.

Consider first $l = 1$. Then $\sum_{i=1}^{m} \exp\left\{-\frac{1}{2}H_{1i}\right\} = 7.6939$. In the same way, when $l = 2$, $\sum \exp\left\{-\frac{1}{2}H_{2i}\right\} = 7.0027$.

The value of the evidence is then

$$V = \frac{\begin{array}{c} 0.0577 \times 8.3189 \times 4.665 \\ \times 10^{-6} \times 0.276 \times 7.237 \end{array}}{\begin{array}{c} 1.29 \times 10^{-5} \times 7.6939 \times 1.29 \\ \times 10^{-5} \times 7.0027 \end{array}} = 498.85.$$

The likelihood ratio correctly supports the proposition according to which the control and recovered fragments originate from the same source.

Consider now the case where the control and recovered fragments are chosen to come from different sources.

Consider the determination of the numerator of the likelihood ratio where the control and recovered fragments are assumed to originate from the same source (i.e. hypothesis H_p is true). The first terms in the numerator that involve a square root of the determinant of the covariance matrix C, and the square root of the inverse of the determinant of a function of covariance matrices U and C are

$$| C |^{1/2} = 0.0577,$$
$$| D_1^{-1} + D_2^{-1} + (h^2 C)^{-1} |^{-1/2} = 4.665 \times 10^{-6}.$$

The smoothing parameter h is estimated as before, so $mh^p = 8.3189$.

The term H_3 has been calculated in Section 7.6.4.1, so that $\exp\left\{-\frac{1}{2}H_3\right\}$ equals 4.234×10^{-18}. The expression

$$\sum_{i=1}^{m} \exp\left\{-\frac{1}{2}H_i\right\}$$

equals 9.377.

Consider the determination of the denominator of the likelihood ratio where the control and recovered fragments are assumed to originate from different sources (i.e. hypothesis H_d is true).

The first terms in the denominator that involve inverses and square roots of inverses of determinants of functions of covariance matrices U and C

are, as before,

$$| D_1^{-1} + (h^2C)^{-1} |^{-1/2} = | D_2^{-1} + (h^2C)^{-1} |^{-1/2} = 1.29 \times 10^{-5}.$$

Note again that $n_1 = n_2$, and $D_1 = D_2$.

Consider first $l = 1$. Then $\sum_{i=1}^{m} \exp\left\{-\frac{1}{2}H_{1i}\right\} = 7.6939$. In the same way, when $l = 2$, $\sum_{i=1}^{m} \exp\left\{-\frac{1}{2}H_{2i}\right\} = 9.4339$.

The value of the evidence is then

$$V = \frac{\begin{array}{c}0.0577 \times 4.665 \times 10^{-6} \times 8.3189 \\ \times 4.233 \times 10^{-18} \times 9.377\end{array}}{\begin{array}{c}1.29 \times 10^{-5} \times 1.29 \times 10^{-5} \\ \times 7.6939 \times 9.4339\end{array}}$$

$$= 7.35 \times 10^{-15}.$$

The likelihood ratio correctly supports the proposition according to which the control and recovered fragments originate from different sources.

The same-source likelihood ratio equals 498.85, the different-source likelihood ratio equals 7.35×10^{-15}. The biggest contributor to this difference in values is the term $\exp\{-\frac{1}{2}H_3\}$, which equals 0.276 for the same-source comparison and 4.233×10^{-18} for the different-source comparison. The term H_3 (7.31) measures the difference between the mean of the control and recovered fragments. If they are dissimilar H_3 is

large and hence $\exp\{-\frac{1}{2}H_3\}$ is small. If they are similar H_3 is small and hence $\exp\{-\frac{1}{2}H_3\}$ is large.

In both these examples the likelihood ratio for different source comparisons is of the order of 10^{-15}. It is proposed that caps of 10^9 and 10^{-9} be placed on likelihood ratios. Jurists will find it hard to believe in values of greater magnitude.

7.6.5 Three-Level Models

So far in this chapter, the discussion has been about two-level models. It is not uncommon for there to be three levels of variation in the data. For example, this may occur when there is measurement error on individual items. This error provides a third level of variation along with variation within and between groups. An example is given in Aitken et al. (2006) that concerns data from a scanning electron microscope coupled with energy dispersive X-ray detectors (SEM-EDX). The instrument provides elemental concentration information from float glass. The three levels of variation are measurement error (relating to the precision of the instrument), within-source error (variation of measurements made on the same object) and between-source variation (variation between measurements made on different objects of the same evidential type). A brief summary of the model is given here. Full details are given in Aitken et al. (2006), where

graphical models are implemented to address the problem of dimensionality.

Let Ω denote a population of p characteristics of items of a particular type. Background data are available of continuous measurements of these characteristics on a random sample of $N = mt$ members (t members from each of m groups) from Ω with $n(\geq 2)$ independent replicate measurements on each of the N members. For example, the data could be the logarithms of the ratio of the concentration of oxygen to each of several other elements in an analysis of float glass; Aitken et al. (2006) consider the logarithmic ratio of oxygen to each of sodium, magnesium, aluminium, silicon, and calcium, so $p = 5$. The background data are denoted as $\mathbf{x}_{ikj} = (x_{ikj1}, \ldots, x_{ikjp})^T; i = 1, \ldots, m; k = 1, \ldots, t; j = 1, \ldots, n;$ with $\bar{\mathbf{x}}_{ik\cdot} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_{ikj}; \bar{\mathbf{x}}_{i\cdot\cdot} = \frac{1}{t}\sum_{k=1}^{t} \bar{\mathbf{x}}_{ik\cdot}$; and $\bar{\mathbf{x}}_{\cdots} = \frac{1}{m}\sum_{i=1}^{m} \bar{\mathbf{x}}_{i\cdot\cdot}$. The data $\mathbf{x}$ are used to estimate various parameters of Ω.

Control data are denoted $\mathbf{y}_1$ and recovered data denoted $\mathbf{y}_2$, as in Section 7.6.2. For control data, denote the number of replicate measurements as n_1 on each of n_c items. For recovered data, denote the number of replicate measurements as n_2 on each of n_s items.

Define a subscript l, which takes one of two values corresponding to whether the data are control ($l = 1$) or recovered ($l = 2$). The control and recovered measurements $\mathbf{y}_1, \mathbf{y}_2$ are vectors with elements $\mathbf{y}_{lkj} = (y_{lkj1}, \ldots, y_{lkjp})^T, j = 1, \ldots, n_l;$

$k = 1, \ldots, n_x; l = 1, 2; x = c$ if $l = 1; x = s$ if $l = 2$. Let $n_0 = (n_1 n_c + n_2 n_s)$, $\bar{\mathbf{y}}_{lk} = \frac{1}{n_l} \sum_{j=1}^{n_l} \mathbf{y}_{lkj}$, $\bar{\mathbf{y}}_l = \frac{1}{n_x} \sum_{k=1}^{n_x} \bar{\mathbf{y}}_{lk}$, and $\bar{\mathbf{y}} = (n_1 n_c \bar{\mathbf{y}}_1 + n_2 n_s \bar{\mathbf{y}}_2)/n_0$.

The model assumes three sources of variation, that of measurement error (replicate measurements on the same item), that between items within the same group (known as within-group variation), and that between groups (known as between-group variation). It is assumed that the variation at all three levels is Normally distributed.

Replication: Denote the mean vector within item k in group i as $\boldsymbol{\theta}_{ik}$ and the covariance matrix of replicate variability as U, constant over all items and groups. Then, given $\boldsymbol{\theta}_{ik}$ and U, the distribution of $\mathbf{X}_{ikj}$ is taken to be Normal:

$$(\mathbf{X}_{ikj} \mid \boldsymbol{\theta}_{ik}, U) \sim N(\boldsymbol{\theta}_{ik}, U);$$
$$i = 1, \ldots, m; \;\; k = 1, \ldots, t; \;\; j = 1, \ldots, n.$$

Within-group: Denote the mean vector within group i by $\boldsymbol{\mu}_i$ and the within-group covariance matrix by V. Then, given $\boldsymbol{\mu}_i$ and V, the distribution of $\boldsymbol{\theta}_{ik}$ is taken to be Normal:

$$(\boldsymbol{\theta}_{ik} \mid \boldsymbol{\mu}_i, V) \sim N(\boldsymbol{\mu}_i, V), ; \;\; i = 1, \ldots, m; \;\; k = 1, \ldots, t.$$

Between-group: Denote the mean vector between groups by $\boldsymbol{\phi}$ and the between-group covariance matrix by W. The distribution of the $\boldsymbol{\mu}_i$, as a measure of between-source variability, is taken to be Normal:

$$(\boldsymbol{\mu}_i \mid \boldsymbol{\phi}, W) \sim N(\boldsymbol{\phi}, W); \;\; i = 1, \ldots, m.$$

Consider the control data $\mathbf{y}_{1kj}, k = 1, \ldots, n_c, j = 1, \ldots, n_1$. Then

$$
\begin{aligned}
(\mathbf{Y}_{1kj} \mid \boldsymbol{\theta}_{i_1 k}, U) &\sim N(\boldsymbol{\theta}_{i_1 k}, U), \\
(\bar{\mathbf{Y}}_{1k} \mid \boldsymbol{\theta}_{i_1 k}, U) &\sim N(\boldsymbol{\theta}_{i_1 k}, n_1^{-1} U), \\
(\bar{\mathbf{Y}}_{1k} \mid \boldsymbol{\mu}_{i_1}, U, V) &\sim N(\boldsymbol{\mu}_{i_1}, n_1^{-1} U + V), \\
(\bar{\mathbf{Y}}_{1} \mid \boldsymbol{\mu}, U, V) &\sim N(\boldsymbol{\mu}, (n_1 n_c)^{-1} U + n_c^{-1} V), \\
(\bar{\mathbf{Y}} \mid \boldsymbol{\phi}, U, V, W) &\sim N(\boldsymbol{\phi}, (n_1 n_c)^{-1} U + n_c^{-1} V + W),
\end{aligned}
$$

where $\boldsymbol{\theta}_{i_1 k}$ is the mean of the replicate measurements on the kth member of the control group. The control group is indicated i_1.

Analogous results follow for the recovered data with i_2 denoting the recovered group so that, for example, $\boldsymbol{\mu}_{i_2}$ replaces $\boldsymbol{\mu}_{i_1}$.

7.6.5.1 Parameter Estimation

The overall mean $\boldsymbol{\phi}$ is estimated by $\bar{\mathbf{x}}_{i..}$, the mean vector over all groups in the background database.

The measurement error (replicate error) covariance matrix U is estimated from the background data $\{\mathbf{x}_{ikj}\}$ by

$$
\widehat{U} = \frac{S_U}{\{mt(n-1)\}}, \tag{7.44}
$$

where $S_U = \sum_{i=1}^{m} \sum_{k=1}^{t} \sum_{j=1}^{n} (\mathbf{x}_{ikj} - \bar{\mathbf{x}}_{ik.})(\mathbf{x}_{ikj} - \bar{\mathbf{x}}_{ik.})^T$.

The within-group covariance matrix V is estimated from the background data $\{\mathbf{x}_{ikj}\}$ by

$$
\widehat{V} = \frac{S_W}{\{m(t-1)\}} - \frac{\widehat{U}}{n}, \tag{7.45}
$$

where $S_W = \sum_{i=1}^{m} \sum_{k=1}^{t} (\bar{\mathbf{x}}_{ik.} - \bar{\mathbf{x}}_{i..})(\bar{\mathbf{x}}_{ik.} - \bar{\mathbf{x}}_{i..})^T$.

The between-group covariance matrix W is estimated from the background data $\{\mathbf{x}_{ikj}\}$ by

$$\widehat{W} = \frac{S_B}{(m-1)} - \frac{\widehat{V}}{t} - \frac{\widehat{U}}{tn}, \qquad (7.46)$$

where $S_B = \sum_{i=1}^{m} (\bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...})(\bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...})^T$.

7.6.5.2 Likelihood Ratio Using a Multivariate Random Effects Model and Assumptions of Normality

The value of the evidence $\mathbf{y}_1$ and $\mathbf{y}_2$ is the ratio of two probability density functions of the form $f(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\phi}, U, V, W)$, one for the numerator, where H_p is assumed true, and one for the denominator, where H_d is assumed true. In the numerator the source means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are assumed equal (to $\boldsymbol{\theta}$, say) but unknown. In the denominator it is assumed that the source means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ need not be equal.

In the numerator denote the probability density function by $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\phi}, U, V, W)$. It is given by

$$\int\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, V) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, W) d\boldsymbol{\mu} d\boldsymbol{\theta},$$

where the four probability density functions are multivariate normal.

In the denominator, the probability density function, denoted $f_1(\mathbf{y}_1, \mathbf{y}_2 \mid \boldsymbol{\phi}, U, V, W)$, is given by

$$\int\int f(\mathbf{y}_1 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, V) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, W) d\boldsymbol{\mu} d\boldsymbol{\theta}$$
$$\times \int\int f(\mathbf{y}_2 \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, V) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, W) d\boldsymbol{\mu} d\boldsymbol{\theta},$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are taken to be independent as the data are assumed to be from different sources.

The value V of the evidence can be shown to be proportional to the ratio of

$$\exp\left(-\frac{1}{2}(H_1 + H_2)\right) \text{ to } \exp\left(-\frac{1}{2}(H_3 + H_4)\right)$$

where

$$H_1 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T \left(\frac{n_1 n_c n_2 n_s U^{-1}}{n_0}\right)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2),$$

$$H_2 = (\bar{\mathbf{y}} - \boldsymbol{\phi})^T \left(\frac{U}{n_0} + (V + W)\right)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\phi}),$$

$$H_3 = (\bar{\mathbf{y}}_1 - \boldsymbol{\phi})^T [(n_1 n_c)^{-1} U + (V + W)]^{-1} (\bar{\mathbf{y}}_1 - \boldsymbol{\phi}),$$

$$H_4 = (\bar{\mathbf{y}}_2 - \boldsymbol{\phi})^T [(n_2 n_s)^{-1} U + (V + W)]^{-1} (\bar{\mathbf{y}}_2 - \boldsymbol{\phi}).$$

The constant of proportionality is a function of the estimates of the population covariance matrices U, V, and W.

The terms H_1, H_2, H_3, and H_4 account for similarity between control and recovered data (H_1) and various measures of rarity: H_2 for overall rarity comparing the overall mean of control and recovered data with the estimate of the overall mean of the background population; H_3 and H_4 for the rarity of the control and recovered data separately.

A full exposition, with a description of the use of graphical models in the reduction of dimensionality, is given in Aitken et al. (2006).

7.7 DISCRIMINATION

Forensic scientists are sometimes faced with the problem of assigning an observation (e.g. an individual, an item, evidence) to one of several populations on the basis of the available measurements of some attributes.

First, consider only two populations, P_1 and P_2. There are n independent measurements $\mathbf{z} = \{z_1, \ldots, z_n\}$, which have come from P_1 or P_2. It is desired to evaluate the strength of the evidence in the form of $\mathbf{z}$ in support of membership of P_1 or P_2.

This problem should be contrasted with the problem of comparison that is prevalent throughout the rest of the book. In a comparison problem, there are control and recovered data. Training data are available from which underlying population parameters may be drawn. The propositions used for the likelihood ratio are the control and recovered data come from the same source (H_p) and the control and recovered data come from different sources (H_d). No statement is made as to what the source may be when H_d is assumed true. In a discrimination problem, there is only one set of evidential data, often denoted $\mathbf{z}$. Training data are available from two populations and population parameters are estimated for both populations. The probability distribution of $\mathbf{Z}$, the corresponding random variable, is known from previous experience for each of the two

populations. Assume each population distribution is parameterised by a known parameter θ_i for population $P_i, i = 1, 2$, with corresponding probability density function $f(\mathbf{z} \mid \theta_i)$, $\theta_i \in \Theta$. The likelihood ratio is then

$$V = \frac{f(\mathbf{z} \mid \theta_1)}{f(\mathbf{z} \mid \theta_2)}$$

and values of $V > 1$ support membership of P_1 for $\mathbf{z}$ and values of $V < 1$ support membership of P_2 for $\mathbf{z}$. This idea can be extended to the situation where there is more than one population so long as prior probabilities for each population may be assigned. Denote the populations $P_1, \ldots, P_k$ with prior probabilities $p_1, \ldots, p_k$ and parameters $\theta_1, \ldots, \theta_k$. Consider the propositions H_p that $\mathbf{z}$ comes from P_1 and H_d that $\mathbf{z}$ comes from one of $P_2, \ldots P_k$. The likelihood ratio is then

$$\frac{f(\mathbf{z} \mid H_p)}{f(\mathbf{z} \mid H_d)} = \frac{f(\mathbf{z} \mid \theta_1)(1 - p_1)}{\sum_{i=2}^{k} f(\mathbf{z} \mid \theta_i) p_i},$$

a continuous analogue of (6.10).

When the θ_i are not known, a marginal probability distribution $f_i(\mathbf{z})$ can be obtained for each population as

$$f_i(\mathbf{z}) = \int_{\Theta} f(\mathbf{z} \mid \theta_i) f(\theta_i) d\theta_l,$$

where $f(\theta_i)$ is a prior probability distribution reflecting available knowledge about θ_i. The

likelihood ratio for discrimination between two populations is then

$$V = \frac{f_1(\mathbf{z})}{f_2(\mathbf{z})}.$$

This is a Bayes' factor (Section 2.3.2).

This can be combined with the prior odds $\Pr(P_1)/\Pr(P_2)$ in favour of P_1 to produce the posterior odds $\Pr(P_1 \mid \mathbf{z})/\Pr(P_2 \mid \mathbf{z})$. The observations can be assigned to population P_1 (P_2) whenever the posterior odds in favour of P_1 is greater (smaller) than 1. It is also possible to include costs in the process of assignation through the use of decision analysis.

7.7.1 Discrete Data

Imagine a scenario where some banknotes (n) are seized on a suspect and some of them (z) after inspection are found to be contaminated with cocaine. These numbers are the evidence E. A typical question a forensic scientist may be called to answer is whether or not the banknotes have been connected with drug trafficking.

Consider a population of banknotes seized during trafficking investigations (P_1), and a population of banknotes in general circulation (P_2). In a simplistic model, the number Z_i of banknotes contaminated in a sample of size n can be mod-

elled by a binomial distribution, $Z_i \sim Bin(n_i, \theta_i)$, $i = 1, 2$, where $\theta_{1(2)}$ denotes the probability that a banknote is contaminated in each of the two populations. It is known that banknotes may be contaminated with drugs in a higher or lower proportion depending on whether they have or have not been involved in drug dealing.

Consider two propositions:

H_p: the banknotes have been involved in drug dealing;

H_d: the banknotes are part of the general circulation.

The value of evidence in a particular case can be determined by considering the ratio of two binomial probabilities and is given by

$$V = \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} = \frac{\binom{n}{z} \theta_1^z (1-\theta_1)^{n-z}}{\binom{n}{z} \theta_2^z (1-\theta_2)^{n-z}}. \qquad (7.47)$$

The proportions θ_1 and θ_2 being typically unknown, a beta prior distribution can be introduced as in Section 4.2, $\theta_i \sim Be(\alpha_i, \beta_i)$, $i = 1, 2$. The marginal probability distribution $f_i(z)$ is a beta-binomial distribution (Section A.2.7) with parameters n, α_i, β_i

$$f_i(z) = \binom{n}{z} \frac{\Gamma(\alpha_i + \beta_i)\Gamma(\alpha_i + z)\Gamma(\beta_i + n - z)}{\Gamma(\alpha_i)\Gamma(\beta_i)\Gamma(\alpha_i + n + \beta_i)}.$$

The likelihood ratio for support for H_p against H_d is then

$$\frac{f_1(z \mid H_p)}{f_2(z \mid H_d)} = \frac{\Gamma(\alpha_1+\beta_1)\Gamma(\alpha_1+z)\Gamma(\beta_1+n-z)\Gamma(\alpha_2)\Gamma(\beta_2)\Gamma(\alpha_2+n+\beta_2)}{\Gamma(\alpha_2+\beta_2)\Gamma(\alpha_2+z)\Gamma(\beta_2+n-z)\Gamma(\alpha_1)\Gamma(\beta_1)\Gamma(\alpha_1+n+\beta_1)}.$$

Consider data concerning numbers of banknotes contaminated with drugs in the two groups, as shown in Table 7.13 (Besson, 2003).

The first group is banknotes seized during drug trafficking investigations, the second group is banknotes in general circulation. Parameters (α_1, β_1) and (α_2, β_2) of the beta prior distributions that have been introduced to model uncertainty about proportions θ_1 and θ_2 of contaminated banknotes in the two populations can be elicited according to the procedure that was described in Section 4.3.1, (4.21) and (4.22). This provides a

Table 7.13 Numbers of banknotes contaminated.

	Number contaminated	Number not contaminated	Total
Notes seized in drug dealing	382	80	462
Notes in general circulation	562	430	992

$Be(381, 80)$ distribution for θ_1, and a $Be(561,430)$ distribution for θ_2. Suppose that a sample of size $n = 100$ is seized and 66 banknotes are found to be contaminated ($z = 66$). The values of the beta-binomial densities can be computed using a statistical software (e.g. R) and the likelihood ratio results

$$\frac{f_1(z \mid H_p)}{f_2(z \mid H_p)} = \frac{\Gamma(942)\Gamma(447)\Gamma(114)\Gamma(561)\Gamma(430)\Gamma(231)}{\Gamma(991)\Gamma(627)\Gamma(464)\Gamma(381)\Gamma(80)\Gamma(401)} = 1/125. \quad (7.48)$$

Note that this is a Bayes' factor as it depends also on the prior parameters and does not simplify to a ratio of likelihoods (see Section 7.9).

The values of the evidence for certain values of n and z are given in Table 7.14.

The results ($n = 100, z = 66$) and ($n = 50, z = 33$) are such that they support H_d. As the ratio z/n increases for fixed n, the strength of the evidence in support of H_p increases. For the same z/n, the larger z and n the stronger the support is in favour of H_p or H_d.

Alternatively, a scientist might opt for a standard likelihood procedure for comparing two propositions in presence of unknown parameters, where the sample proportions of the parameters for each hypothesis, that is, $\hat{\theta}_1$ and $\hat{\theta}_2$, are substituted into the associated terms in the numerator and

Table 7.14 The value of evidence in comparing two binomial proportions for different values of sample size n and number of 'successes' z, where the samples are banknotes and a 'success' is a contaminated banknote.

Sample size	Number contaminated	Proportion	Evidential value
n	z	z/n	V
100	66	0.66	1/123
100	72	0.72	5.65
100	76	0.76	516
50	33	0.66	1/15
50	36	0.72	2.25
50	38	0.76	25.23

Population parameters θ_1 and θ_2 are modelled by a beta prior distribution, $Be(381, 80)$ and $Be(561,430)$, respectively.

denominator in (7.47) as

$$V = \frac{\binom{n}{z} \hat{\theta}_1^z (1 - \hat{\theta}_1)^{(n-z)}}{\binom{n}{z} \hat{\theta}_2^z (1 - \hat{\theta}_2)^{(n-z)}}.$$

If H_p is true then the proportion of banknotes contaminated with drugs is estimated to be $\hat{\theta}_1 = 382/462 = 0.83$, from the data in Table 7.13 and, similarly, if H_d is true then the proportion of banknotes contaminated with drugs is estimated to be $\hat{\theta}_2 = 562/992 = 0.57$. The value V of the evidence E in favour of the

proposition that the banknotes have been involved in drug dealing is then

$$V = \frac{\binom{n}{z} 0.83^z(1-0.83)^{(n-z)}}{\binom{n}{z} 0.57^z(1-0.57)^{(n-z)}}$$
$$= \left(\frac{0.83}{0.57}\right)^z \left(\frac{0.17}{0.43}\right)^{(n-z)} = 1/854.$$

The use of the binomial distribution may be questioned here. One of the modelling assumptions for a binomial distribution is that all members of the sample have a constant probability of 'success', independent of other members of the sample. Models that allow for dependence amongst members of the sample are beyond the scope of this book; see, for example, Wilson et al. (2014, 2015).

7.7.2 Continuous Data

As a theoretical example with two continuous populations, assume $\theta_i = (\mu_i, \sigma^2)(i = 1, 2)$ and a univariate random variable $Z \sim N(\mu_1, \sigma^2)$ for population P_1 and $Z \sim N(\mu_2, \sigma^2)$ for population P_2. Let H_p be the proposition that Z is a member of population P_1 and H_d be the proposition that Z is a member of population P_2. Consider a set $\mathbf{z} = \{z_1, \ldots, z_n\}$ of n observations. Then the

probability density function for $\mathbf{z}$ in population P_i is

$$f_i(\mathbf{z}) = \prod_{j=1}^{n} f(z_j \mid \mu_l, \sigma^2)$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} \exp\,(z_j - \mu_i)^2 \right\}.$$

The likelihood ratio for support for H_p against H_d is then

$$\frac{\prod_{j=1}^{n} f(z_j \mid H_p, \mu_1, \sigma^2)}{\prod_{j=1}^{n} f(z_j \mid H_d, \mu_2, \sigma^2)}$$

$$= \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{j=1}^{n} (z_j - \mu_1)^2 - \sum_{j=1}^{n} (z_j - \mu_2)^2 \right] \right\}$$

$$= \exp\left\{ -\frac{n(\mu_1 - \mu_2)}{\sigma^2} \left[\frac{1}{2}(\mu_1 + \mu_2) - \bar{z} \right] \right\}.$$

Assume $\mu_1 > \mu_2$. Then the likelihood ratio is less than 1, and hence supportive of the proposition that $\mathbf{z}$ comes from population P_2 with mean μ_2, if $\bar{z} < (\mu_1 + \mu_2)/2$, i.e. if $\bar{z}$ is less than the mean of the two proposition means, an intuitively reasonable result given the precision (reciprocal of the variance) with which the data are measured in each population is the same.

An example with unequal population variances is given in Taroni et al. (2010). A scientist is interested in evaluating the measurement of colour dye concentration in ecstasy tablets.

A comparison is to be made of the measurement z on a tablet of unknown origin and measurements on a consignment (C) of tablets for which laboratory analysis has revealed the presence of a certain kind of colour dye. One proposition (H_p) for discrimination is that the tablet comes from C; the other proposition (H_d) is that it comes from a population of unrelated cases, denoted P. The concentration of colour dye is a continuous measurement (expressed as a percentage) with a Normal distribution. For tablets from C, the distribution is assumed to be $N(\mu_c, \sigma_c^2)$. For tablets from P, the distribution is assumed to be $N(\mu_p, \sigma_p^2)$. The likelihood ratio for measurement z is then the ratio of two Normal density functions

$$V = \frac{f(z \mid H_p, \mu_c, \sigma_c^2)}{f(z \mid H_d, \mu_p, \sigma_p^2)} = \frac{\sigma_p \exp\left\{-\frac{1}{2}\left(\frac{z-\mu_c}{\sigma_c}\right)^2\right\}}{\sigma_c \exp\left\{-\frac{1}{2}\left(\frac{z-\mu_p}{\sigma_p}\right)^2\right\}}.$$

An example is given using data from Goldmann et al. (2004). A tablet is analysed and the colour concentration measured equals 0.155%. Distributions of competing propositions for populations C and P are taken to be $N(0.14, 0.01^2)$ and $N(0.30, 0.06^2)$, respectively. The likelihood ratio for evaluating a link to C is then

$$V = \frac{f(0.155 \mid H_p, 0.14, 0.01^2)}{f(0.155 \mid H_d, 0.30, 0.06^2)} \simeq 36.$$

The observed colour dye concentration in the incriminated tablet may thus be said to be approximately 36 times more likely if it is linked to C than to P.

Consider the case where the mean μ_i of the Normal distribution is unknown and the variance σ^2 is assumed common and known to both populations. Suppose now there is prior information about the probability distribution of μ for each of the two populations so that $\mu_i \sim N(\eta_i, \tau^2), i = 1, 2$. The distribution of Z can be shown to be $N(\eta_i, \sigma^2 + \tau^2)$ (see Section 7.4.1) and the likelihood ratio can then be determined as before but with an increased variance to allow for the uncertainty surrounding μ_1 and μ_2.

Further examples of discrimination are given in Taroni et al. (2010) by taking advantage of a Bayesian decision perspective. In particular, an example is given where the variance as well as the mean is unknown. The resulting probability distribution for the measurement is a Student's t-distribution. This example is illustrated with the examination of skeletal remains. The two populations under consideration are those of males and of females. Analysis and measurements of the sacral base (*basis osseus sacri*) is considered a good determinant of sex (Benazzi et al., 2009). Another example is that of scenarios in microbial forensic science for the identification of source cultures generated by a pathogen (Lindgren et al., 2019).

7.7.3 Autocorrelated Data

The methodology for discrimination described so far has assumed the observations are independent. An example with univariate data where this assumption does not hold is that of the quantity of cocaine on banknotes. Consider a bundle of banknotes. The quantity of cocaine on a banknote is associated positively with the quantity of cocaine on its immediate neighbour. Further details are available in Wilson et al. (2014, 2015).

The banknotes on which the quantities of cocaine are measured in order to determine the value of the evidence (V) are provided by law enforcement agencies to an analytical chemistry laboratory. The laboratory measures the quantity of cocaine either on all of the banknotes or on a subset of the banknotes and is interested in the support provided by the results for one or other of the following propositions:

H_p: the banknotes are associated with a person who is associated with a criminal activity involving cocaine;

H_d: the banknotes are associated with a person who is not associated with a criminal activity involving cocaine.

The data used for the analysis are the logarithms of the peak areas (where the peaks are obtained using a mass spectrometer) corresponding to cocaine on a set of banknotes. Training data are

available, first, from banknotes from cases that went to trial and in which the defendant was convicted (either by trial or through a plea of guilty) of a crime involving cocaine and, second, from banknotes associated with general or background circulation.

There has been some suggestion of regional variation in the quantity of cocaine on banknotes in general circulation. However, Aitken et al. (2017) shows that there is no meaningful variation of the quantities of cocaine on banknotes in general circulation in England and Wales.

7.7.4 Multivariate Data

Often, multivariate continuous data are available for discrimination between two or more sources. Evidence such as handwritten or printed characters in a questioned document, or fragments of glass recovered at a crime scene, or a drug sample can be characterised by more than one variable. A statistical model for the evaluation of evidence through the computation of a likelihood ratio for multivariate data has been proposed by Aitken and Lucy (2004) in the context of the elemental composition of glass data, and by Bozza et al. (2008) in the context of handwritten questioned documents, both for comparison problems. An example for discrimination is given here.

Consider a scenario involving questioned documents that have been printed by one of two printers

(P_1 and P_2). The two printers are the subject of the two propositions for discrimination:

H_p: the questioned document has been printed with printer P_1 (e.g. *Canon* model ir400);

H_d: the questioned document has been printed with printer P_2 (e.g. *HP* model 41).

Note that there is not necessarily an association in this case between printer and prosecution and defence.

There are several (p) variables that may be measured on each printed character (e.g. the area (Mazzella and Marquis, 2007)). The background data consist of $n_i, i = 1, 2$, measurements of these variables on characters printed with each of the two printers, and are denoted as $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T, \quad i = 1, 2, \quad j = 1, \ldots, n_i$. The procedure outlined earlier for two univariate Normal populations is extended to the case of multivariate data. Once new printed documents of unknown origin become available, the problem becomes one of discrimination of their source between the two multivariate Normal populations, corresponding to the two printers, say, $N(\boldsymbol{\theta}_1, \Sigma)$ and $N(\boldsymbol{\theta}_2, \Sigma)$, where $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{ip})^T$ is the vector of means of the i-th population, $i = 1, 2$, and Σ is the matrix of variances and covariances of each population, assumed to be the same in each population.

Denote the recovered measurements to be classified by $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T$, where $\mathbf{z}_j = (z_{j1}, \ldots, z_{jp})^T$,

$j = 1, \ldots, n$. The probability density of a multivariate Normal variable $\mathbf{Z}$ (Appendix A.3.9) $N(\boldsymbol{\theta}, \Sigma)$ is

$$f(\mathbf{z} \mid \boldsymbol{\theta}, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\theta})\right].$$

If the populations are exactly known (i.e. the mean vectors and the covariance matrix are known), the value V of evidence is given by

$$\begin{aligned} V &= \frac{f(\mathbf{z} \mid \boldsymbol{\theta}_1, \Sigma)}{f(\mathbf{z} \mid \boldsymbol{\theta}_2, \Sigma)} \\ &= \frac{\prod_{j=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{z}_j-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\mathbf{z}_j-\boldsymbol{\theta}_1)\right]}{\prod_{j=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{z}_j-\boldsymbol{\theta}_2)^T\Sigma^{-1}(\mathbf{z}_j-\boldsymbol{\theta}_2)\right]} \\ &= \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^{n}(\mathbf{z}_j-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\mathbf{z}_j-\boldsymbol{\theta}_1) \right.\right. \\ &\quad \left.\left. -\sum_{j=1}^{n}(\mathbf{z}_j-\boldsymbol{\theta}_2)^T\Sigma^{-1}(\mathbf{z}_j-\boldsymbol{\theta}_2)\right]\right\}. \end{aligned} \tag{7.49}$$

Generally, population distributions will not be completely known. A simple criterion for discrimination consists in estimating $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and Σ from the background data $\{\mathbf{x}_{ij}\}$ and substituting them into (7.49). The mean vector $\boldsymbol{\theta}_i$ can be estimated by

$$\bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}, \qquad i = 1, 2.$$

The covariance matrix Σ can be estimated by

$$S = \frac{1}{n_1 + n_2 - 2}\left[\sum_{j=1}^{n_1}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)^T + \sum_{j=1}^{n_2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)^T\right]. \qquad (7.50)$$

Consider an example with two variables, the area and the box-ratio of the letter 'a', thus $p = 2$. Two ($n = 2$) characters of type 'a' are measured from a questioned document. The document might have been printed by a *Canon* model *ir*400 (printer P_1, hypothesis H_p), or by an *HP* model 4*l* (printer P_2, hypothesis H_d). The mean vectors $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$, which are used as estimates of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, are $\bar{\mathbf{x}}_1 = (462\,550, 1.24)^T$ and $\bar{\mathbf{x}}_2 = (430\,350, 1.32)^T$. The covariance matrix Σ is estimated by S, from (7.50), and is

$$S = \begin{pmatrix} 2.37\text{e}+0.8 & -2.11\text{e}+0.2 \\ -2.11\text{e}+0.2 & 9.98\text{e}-04 \end{pmatrix}.$$

Substituting $\hat{\boldsymbol{\theta}}_i = \bar{\mathbf{x}}_i$, $i = 1, 2$, and $\hat{\Sigma} = S$ in (7.49) gives

$$V = \frac{f(\mathbf{z} \mid \bar{\mathbf{x}}_1, S)}{f(\mathbf{z} \mid \bar{\mathbf{x}}_2, S)} = 6\,074.$$

The available measurements from the questioned document are approximately 6000 times more likely if the questioned document has been printed with printer P_1.

A Bayesian framework for multivariate discrimination would require the introduction of a prior distribution for the unknown parameters, and it is briefly sketched later. Consider the simplest case where the prior distribution of the mean vectors $\boldsymbol{\theta}_i$ is taken to be Normal, say, $\boldsymbol{\theta}_i \sim N(\boldsymbol{\mu}_i, C)$, $i = 1, 2$. The covariance matrix Σ, conversely, is estimated from the available background data.

The marginal distribution under H_p (the questioned document has been printed with printer P_1), $f(\mathbf{z} \mid \boldsymbol{\mu}_1, C, \Sigma, H_p)$, is given by

$$\int f(\mathbf{z} \mid \boldsymbol{\theta}_1, \Sigma) f(\boldsymbol{\theta}_1 \mid \boldsymbol{\mu}, C) d\boldsymbol{\theta}_1 = \int \prod_{j=1}^{n} |2\pi|^{-p/2} |\Sigma|^{-1/2}$$

$$\exp\left\{-\frac{1}{2}(\mathbf{z}_j - \boldsymbol{\theta}_1)^T \Sigma^{-1} (\mathbf{z}_j - \boldsymbol{\theta}_1)\right\} \times |2\pi|^{-p/2} |C|^{-1/2}$$

$$\times \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\mu}_1)^T C^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\mu}_1)\right\} d\boldsymbol{\theta}_1$$

and can be shown to be equal to

$$f(\mathbf{z} \mid \boldsymbol{\mu}_1, \Sigma, C, H_p)$$

$$= |2\pi\Sigma|^{-n/2} |2\pi C|^{-1/2} |2\pi(n\Sigma^{-1} + C^{-1})^{-1}|^{1/2}$$

$$\exp\left\{-\frac{1}{2}\left[\operatorname{tr}(S\Sigma^{-1}) + (\bar{\mathbf{z}} - \boldsymbol{\mu}_1)^T \left(\frac{1}{n}\Sigma + C\right)^{-1} (\bar{\mathbf{z}} - \boldsymbol{\mu}_1)\right]\right\},$$

where $\bar{\mathbf{z}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{z}_j$, $S = \sum_{j=1}^{n} (\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})^T$.

In the same way the marginal distribution under H_d (the questioned document has been printed with printer 2) is given by $f(\mathbf{y} \mid \boldsymbol{\mu}_2, \Sigma, C, H_d)$.

The value of the evidence is given by

$$V = \frac{f(\mathbf{z} \mid \boldsymbol{\mu}_1, \Sigma, C, H_p)}{f(\mathbf{z} \mid \boldsymbol{\mu}_2, \Sigma, C, H_d)}. \qquad (7.51)$$

In the same way, it is possible to address the case where a prior distribution is considered for both the mean vector and the covariance matrix. However, the elicitation of a prior distribution for the parameters of a multivariate Normal distribution, in particular for the covariance matrix, can be a difficult problem and it will not be pursued anymore. Some suggestions are given in O'Hagan et al. (2006).

Further examples of discrimination problems for univariate and multivariate data may be found in Taroni et al. (2010) and Zadora et al. (2014). It is possible to use kernel density estimation for discrimination if a Normal distribution is deemed inappropriate; see Aitken (1986) and Peabody et al. (1983).

7.7.5 Cut-Offs and Legal Thresholds

Many analytical branches of forensic science, e.g. forensic toxicology, rely on what are called *cut-offs*. These are numerical values against which measurements made on questioned items are compared in order to support an interpretation or conclusion in a forensic toxicological assessment regarding a person of interest. Examples for sets of results are concentrations of toxic or controlled substances in blood or of target

substances (e.g. metabolites) in hair. Examples where these analyses are of wide interest are workplace safety, child custody cases, suspected doping cases in sport and ink dating in forensic document examination. A sample from a person for whom it is not known whether they fall in one or other category of individuals is analysed. The comparison of the measured value against the cut-off provides discriminatory information.

The idea of discrimination for a so-called forensic cut-off in toxicology is critically analysed and discussed in Biedermann et al. (2018). The examples discussed are those of alcohol markers in hair, used widely by forensic toxicologists to reach conclusions regarding the drinking behaviour of individuals. A cut-off is, however, incompatible with current evaluative guidelines (e.g. ENFSI (2015)). A cut-off does not define an offence. The results are advisory.

In the context of hair analysis, the amount of ethyl glucoronide (EtG) could be used to define a cut-off to assist in answering questions about the habit of alcohol consumption in a person of interest, e.g. is the habit one of excessive alcohol consumption? The legal question focuses on the characteristics (drinking behaviour) of the person of interest but there is no legally prescribed limit for a particular substance, nor is there any mention of a particular type of biological tissue (e.g. hair).

As an example of the interpretation of a forensic cut-off, consider the case of a man suspected

of chronic excessive alcohol consumption. The forensic medical diagnosis of excessive alcohol consumption may be necessary, for example, for reasons of safety in the workplace. Hair of the person of interest was analysed following accepted analytical procedures. The analysis of hair revealed a concentration of 28 pg/mg in the proximal segment up to 6 cm. The Society of Hair Testing (SoHT) has declared that a 'concentration of greater than 30 pg/mg EtG in the proximal scalp hair up to 6 cm strongly suggests chronic excessive alcohol consumption'.[2] Thus a concentration of 28 pg/mg would lead to the conclusion that there is no 'strong suggestion' for chronic excessive alcohol consumption. The perspective based on the application of a cut-off limits the analyst to such a conclusion, though this might lead to the fallacious perception that such a comparison of the results with the cut-off value is sufficient to form a conclusion about a particular proposition. Note that no alternative proposition is specified, and so no guidance is provided on what to do if results are below the cut-off. However, consider the principle that evidence should be evaluated with respect to two competing propositions. The use of a cut-off begs the question of what probative value the particular result of 28 pg/mg has with respect to competing propositions of interest.

[2] 2016 Concensus for the use of alcohol markers in hair for the assessment of both abstinence and chronic excessive alcohol consumption (http://soht.org/images/pdf/Revision%202016_Alcoholmarkers.pdf)

Generally, a value of 28 pg/mg is something less typically found among non-heavy drinkers. In the case here, suppose that the two competing propositions of interest are: 'The person of interest is a chronic excessive drinker', and 'The person of interest is a low-risk (i.e. social) drinker'. A different approach to that of the SoHT is taken by Kharbouche et al. (2012). These authors define 'at-risk' drinkers to be those who consume more than 30 g of alcohol per day and 'low-risk male drinkers' to be those who declare, according to a Daily Alcohol Self-Monitoring log (DASM log), that they drink less than 30 g/d. These alternative definitions enable a likelihood ratio to be developed for the EtG measurements. Kharbouche et al. (2012) study the EtG levels for people in each of these groups. Probability distributions of EtG levels for each of the two groups can then be determined. Figure 7.8(a) illustrates EtG concentrations measured in the hair of 14 male low-risk drinkers and 28 male heavy drinkers. Figure 7.8(b) illustrates the same data limited to concentrations below 100 pg/mg EtG (Kharbouche et al., 2012). Ideally, probability densities for the distribution of EtG under each of H_p and H_d would be determined from the data. The ratio of the density functions at the observed value of 28 pg/mg would then provide a likelihood ratio with an interpretation that the observation (28 pg/mg) is so many times more likely if the person of interest is a chronic excessive drinker than if he is a social drinker. However, the data are sparse and only a more

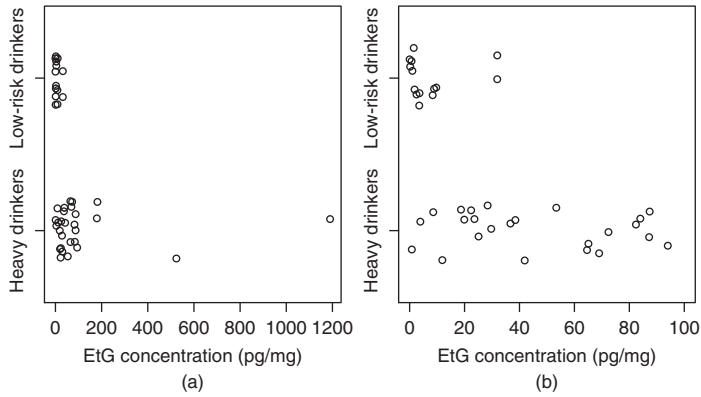

Figure 7.8 (a) EtG concentrations measured in the hair of 14 male low-risk drinkers and 28 heavy drinkers. (b) Representation of the same data limited to concentrations below 100 pg/mg EtG. Note that within each of the two categories displayed on the *y*-axis, the vertical scattering of the values has no particular meaning other than serving the purpose of visually separating values with the same or very similar values on the *x*-axis (EtG concentrations).

general conclusion may be drawn. For example, there are 7 out of 28 monitored heavy drinkers with values in the range 20–40 pg/mg. Thus a probability of 0.25 (7/28) may be assigned to the stated range for heavy drinkers. In contrast there are only 2 out of 14 monitored social drinkers with values in the range 20–40 pg/mg. Thus a probability of approximately 0.14 (2/14) may be assigned to the stated range for social drinkers. The likelihood ratio is then (7/28) divided by (2/14) or $(7 \times 14)/(28 \times 2) = 1.75$. Thus the data are very approximately twice as probable if the person of interest is a chronic excessive drinker

than if he is a social drinker, and this supports the proposition 'heavy drinker' versus the proposition 'social drinker' despite the result being below the cut-off of 30 pg/mg. This result is not an expression of the probative value of the particular result of 28 pg/mg. It is a more coarse description of the findings in terms of a value within a particular range of values, namely, 20–40 pg/mg.

For further details of the development of likelihood ratios for alcohol consumption and the comparison with an approach based on cut-offs, see Biedermann et al. (2018) and Alladio et al. (2017).

It must be observed that a scientific cut-off is effectively not needed when the question of interest to a recipient of expert information relates to discrimination. In contrast, numerical legal thresholds are limiting values that serve the purpose of defining situations that constitute offences. A question that is commonly encountered by toxicologists and forensic scientists in general is whether a target substance is measurable in some examined material and, if so, whether the quantity of the substance of interest exceeds a particular threshold defined by law. For an analysis and discussion, see e.g. Taroni et al. (2014b).

From the perspective of forensic interpretation, numerical legal thresholds can be used as a basis for the definition of propositions of interest. Propositions can be defined from case circumstances and agreed key issues and may reflect the positions taken by adversarial parties at trial.

For example, in a context where the law defines the highest admissible quantity of a target (e.g. toxic) substance in the blood or urine of a person of interest, the propositions of interest become whether or not the person of interest is below or above the legally prescribed limit.

An example is that of drink driving where there is a legal limit. A driver who is found to be above the limit is guilty of an offence. However, there is measurement error and hence random variation in the instrument used to determine the level of alcohol present in the blood, say. Suppose the limit is θ_0. It is of interest to know if the blood alcohol level θ for the driver is above θ_0. Let x be the measured level and assume this has a probability distribution $f(x \mid \theta)$ dependent on the blood level of alcohol. The probability of interest is $\Pr(\theta > \theta_0 \mid x)$. If this probability is sufficiently high then the driver can be charged with drink driving.

Let Θ_p be the set $\{\theta > \theta_0\}$ and Θ_d be the set $\{\theta < \theta_0\}$. There are two propositions:

$H_p : \theta \in \Theta_p$;

$H_d : \theta \in \Theta_d$.

The prior odds are

$$\frac{\int_{\Theta_p} f(\theta)d\theta}{\int_{\Theta_d} f(\theta)d\theta},$$

where $f(\theta)$ is a prior probability distribution reflecting available knowledge about θ.

The posterior odds are

$$\frac{\int_{\Theta_p} f(x \mid \theta) f(\theta) d\theta}{\int_{\Theta_d} f(x \mid \theta) f(\theta) d\theta}.$$

The ratio of the posterior odds to the prior odds provides Bayes' factor. It will be shown in Section 7.9 that this does not simplify to a likelihood ratio.

7.8 SCORE-BASED MODELS

Models for comparison and for discrimination that use the original data are known as feature-based models. The models discussed in Sections 7.6 and 7.7 are all feature-based. Feature-based models described in Section 7.6 compare the probability of observing the evidence given that the evidential samples (control and recovered) measured, and compared, come from the same source as or come from different sources. In contrast, what are known as score-based models and which are considered in the current section compare the probability of observing the pairwise similarity between two samples (control and recovered) given that they come from the same source with the probability of the pairwise similarity given that the samples come from different sources.

Consider the problem of comparison of sources with a p-dimensional control measurement $\mathbf{x} = (x_1, \ldots, x_p)^T$ and a p-dimensional recovered

measurement $\mathbf{y} = (y_1, \ldots, y_p)^T$. For those occasions when a feature-based model is not tractable (e.g. multidimensional binary data, Aitken and Huang (2017)), the distance $d(\mathbf{x}, \mathbf{y})$, known as a *score* can be used instead. A score can be defined as a metric that summarizes the results of a forensic comparison of two items in terms of a univariate statistic, such as a measure of similarity or difference (distance). There are various distance measures that may be used. Three examples are

- Euclidean: $d = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2}$;
- Manhattan: $d = \sum_{i=1}^{p} \mid x_i - y_i \mid$;
- Pearson correlation distance: $100(1 - r)/2$ with

$$r = \frac{\sum_{i=1}^{p}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{p} (x_i - \bar{x})^2 \sum_{i=1}^{p} (y_i - \bar{y})^2}}.$$

For multiple control and recovered data $\mathbf{x}_i, i = 1, \ldots, m$ and $\mathbf{y}_i, i = 1, \ldots, n$, respectively, pairwise score measurements or means can be used.

The value of the evidence is then

$$V = \frac{f(d(\mathbf{x}, \mathbf{y}) \mid H_p, I)}{f(d(\mathbf{x}, \mathbf{y}) \mid H_d, I)}.$$

Inference may then continue as before but using the score, which is univariate, as the statistic of interest. Score-based approaches do not require the distributional assumptions needed to fit feature

based models (e.g. within-source normality) but still require a function to be chosen to model the probability distribution of the score.

For the calculation of score-based likelihood ratios, distributions of scores of same-source comparisons and of different-source comparisons are required. In order to estimate these a training set is required. Determination of the same-source distribution can be made by comparing measurements **x** of the known feature and the features of items known to originate from the same source as **x**, except with **x** itself for which the distance is zero. For the different-source distribution, a challenging task consists in choosing an appropriate training set for measurements comparison. In particular, it can be proposed a (*i*) *source-anchored* approach, where the features **x** of the control items are compared with the features selected randomly from the relevant population; (*ii*) a *trace-anchored* approach, where the feature **y** of the recovered item are compared with the features selected randomly from the relevant population; and (*iii*) a *non-anchored* approach, where features of the items are taken from randomly selected sources from the relevant population. See Hepler et al. (2012) for a discussion.

These results may then be used to estimate the distributions of same-source and between-source comparisons. The distributions can be represented initially by histograms. They may then be smoothed with a kernel density estimator or

an appropriate parametric distribution. The chosen distribution functions, one for same-source comparisons and one for different-source comparisons, can then be used to determine the density calculation (height of the density curve) of the evidence score for both distributions and hence calculate a likelihood ratio.

Score-based approaches have been used for handwriting (Hepler et al., 2012), forensic MDMA comparison (Bolck et al. (2015), speech recognition (Gonzalez-Rodriguez et al. (2006); Brümmer and Du Preez (2006); Morrison (2011)), fingerprints (Egli et al. (2007); Gonzalez-Rodriguez et al. (2005); Leegwater et al. (2017)), signatures (Chen et al., 2019), marks left on gun cartridges (Riva (2011); Riva and Champod (2014); Riva et al. (2017)), and in Morrison and Enzinger (2018) for different branches of forensic science. A review of different approaches proposed in the literature is presented in Jacquet and Champod (2020).

A comparison of the performances of score-based and features-based likelihood ratios for forensic MDMA comparisons is given by Bolck et al. (2015). Feature-based models have the valuable benefit compared with the score-based models of preserving the original data dimensionality with no loss of information. Moreover, with feature-based models, rarity and similarity of the features relate directly to the magnitude of the likelihood ratio. Conversely, feature-based models can be difficult to implement because of the limited population samples to estimate model parameters.

Further comments are given by Neumann and Saunders (2015) and Neumann and Ausdemore (2019).

7.8.1 Example

An experiment was conducted to investigate the evidential value of striation marks in screwdrivers (Petraco et al., 2012). The striation patterns made by each of nine screwdrivers were recorded. Distances of each line or groove from the left edge of each striation pattern were measured to the nearest 0.05 mm. Each striation pattern was no more than 7 mm wide. For each pattern, a list of 140 pieces of information (7/0.05 mm slots) was created. Each piece of information is a 1 or 0. A 1 is recorded in a slot on the list if a line or groove were present or spans the slot. A 0 is recorded otherwise. The sample space, represented as B^{140}, for the data is then a vector of 140 1's and 0's, denoting the presence and absence of a mark, respectively.

In another case, the evidence E would be the striated marks made by a screwdriver presented in the form of a vector in $B^{140} = \{0, 1\}^{140}$. The control evidence, $\mathbf{x}$, is the vector of marks for which the source (screwdriver) is known. This could be a screwdriver found in the possession of a person of interest, for example. The recovered evidence, $\mathbf{y}$, is evidence for which the source is not known. This could be a set of striation marks found at the scene of a crime. These marks could have been made

by the screwdriver found in the possession of the PoI; this would be the prosecution's proposition, H_p. Alternatively, these marks could have been made by some other screwdriver; this would be the defence proposition, H_d.

Training data are available (Petraco et al., 2012) from the nine screwdrivers with (8, 6, 9, 8, 9, 9, 8, 9, 9) replicates for each screwdriver. Let $\mathbf{x}$ and $\mathbf{y}$ be two sets of binary measurements in B^{140}, not necessarily equal to any vector in the training set. The distance $d(\mathbf{x}, \mathbf{y})$ between them is defined as

$$d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y}) = \sum_{i=1}^{140} (x_i - y_i)^2;$$
$$\{x_i, y_i\} \in \{0, 1\}.$$

The result is a number between 0 and 140, inclusive. There are 279 within-group distances and 2496 between-group distances. The distributions of these distances can be estimated with a kernel density function. Histograms and fitted density functions for within-group and between-group distances are shown in Figures 7.9 and 7.10, respectively.

The problem of the evaluation of evidence for a vector consisting of 140 1's and 0's has been transformed into one for continuous data, treating the 141 possible distances 0, ..., 140 as a continuous variable. Most of the within-group distances are below 40, whereas the between-group distances cluster in the interval 30–70. The likelihood ratio for a particular distance is the ratio of the height

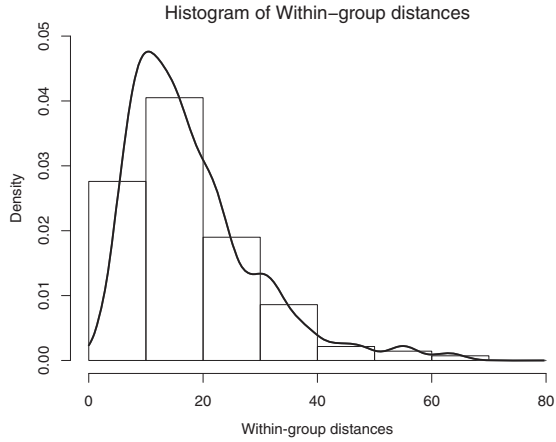

Figure 7.9 Histogram and kernel density estimate for 279 within-group distances for 140 striation marks from nine screwdrivers (Source: Modified from Aitken and Huang, 2017).

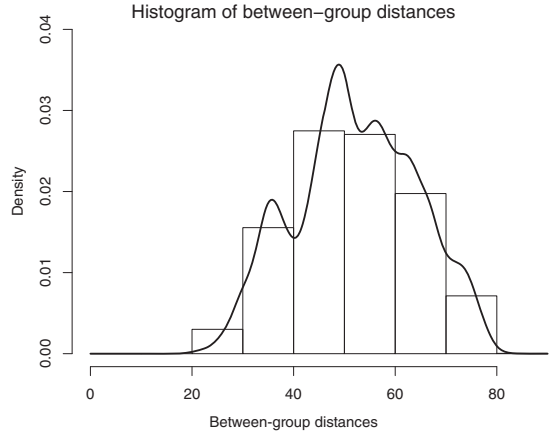

Figure 7.10 Histogram and kernel density estimate for 2496 between-group distances for 140 striation marks from nine screwdrivers (Aitken and Huang, 2017).

of the within-group curve to the height of the between-group curve at that particular distance. As an example of the approach in practice, the distance between the first two sets of striation marks from the first screwdriver is 11 and the likelihood ratio is 2×10^{10}.

The validity of large values of the likelihood ratio, such as the one just given, has to be questioned. Scientists who have more intimate knowledge of the theories and models adopted and data used, should appreciate the limitations of the theories and models and also the limitations of the possible conclusions. The robustness of an adopted model used for the calculation of a likelihood ratio should be empirically supported. Methods to deal with those aspects are presented in Chapter 8. DNA scientists are also faced with this problem. Extremely large likelihood ratio values can be (theoretically) obtained. Scientists are conscious that such large values invoke independence assumptions with a scale of robustness that cannot be demonstrated empirically given the size of currently available DNA profiles databases. Moreover, scientists are also conscious of problems related to the conceptualization of such large values and their relationship with real life situations. A brief discussion of this is presented in Section 6.1.5.

7.9 BAYES' FACTOR AND LIKELIHOOD RATIO (CONT.)

In the current chapter several instances for the determination of a likelihood ratio have been

addressed for evaluative and discriminative purposes, for univariate and multivariate data, and for feature-based and score-based approaches. As already mentioned in Section 2.3.2, the likelihood ratio is in fact a Bayes' factor (BF), though it is generally referred to in forensic science applications as a likelihood ratio (LR). The distinction between a BF and a LR has been discussed in Section 2.3.2 for discrete events, and will now be extended to encompass continuous quantities that have been modelled in this chapter (e.g. population means).

Consider an unknown quantity X and suppose $f(x \mid \theta)$ is a suitable probability model for X, where the unknown parameter θ belongs to the parameter set Θ. Suppose also that the parameter set is partitioned into two non-overlapping sets Θ_p and Θ_d. A question that may be of interest is whether the true but unknown value of the parameter θ belongs to Θ_p, or to Θ_d, that is, to test the hypothesis[3] $H_p : \theta \in \Theta_p$, against the alternative hypothesis $H_d : \theta \in \Theta_d$. A hypothesis is called *simple* if there is only one possible value for the unknown parameter, say, $\Theta_p = \{\theta_0\}$; if a hypothesis is not simple it is called *composite*. Let $\pi_p = \Pr(\theta \in \Theta_p)$ and $\pi_d = \Pr(\theta \in \Theta_d)$ denote one's prior probabilities for the competing hypotheses.

First, consider the case where competing hypotheses are simple, say, $H_p : \theta = \theta_0$ versus $H_d : \theta = \theta_1$. The parameter sets in this case

[3]The term 'hypothesis' is used here instead of 'proposition'. This is for compatibility with the corresponding mathematical statistical exposition.

are $\Theta_p = \{\theta_0\}$ and $\Theta_d = \{\theta_1\}$. Denote the prior probabilities by $\pi_p = \Pr(\theta = \theta_0)$ and $\pi_d = \Pr(\theta = \theta_1)$. If $\pi_p + \pi_d = 1$, the ratio π_p/π_d of the prior probabilities of the competing hypotheses is the prior odds of H_p against H_d.

Suppose a set of observations $x = (x_1, ..., x_n)$ is available. Denote by α_p the posterior probability of the hypothesis H_p given the data and prior probabilities. It can be computed with an application of Bayes' theorem:

$$\alpha_p = \Pr(\theta = \theta_0 \mid x) = \frac{f(x \mid \theta_0)\pi_p}{f(x \mid \theta_0)\pi_p + f(x \mid \theta_1)\pi_d}.$$

The posterior probability α_d of the alternative hypothesis is computed analogously by

$$\alpha_d = \Pr(\theta = \theta_1 \mid x) = \frac{f(x \mid \theta_1)\pi_d}{f(x \mid \theta_0)\pi_p + f(x \mid \theta_1)\pi_d}.$$

The ratio α_p/α_d of the posterior probabilities is the posterior odds of H_p against H_d and is given by

$$\frac{\alpha_p}{\alpha_d} = \frac{f(x \mid \theta_0)\pi_p}{f(x \mid \theta_1)\pi_d}. \tag{7.52}$$

Given that the BF is the ratio between posterior odds to prior odds, it can be shown that

$$\text{BF} = \frac{f(x \mid \theta_0)\pi_p}{f(x \mid \theta_1)\pi_d} \times \frac{\pi_d}{\pi_p} = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)} = \text{LR}.$$

In the case of testing a simple hypothesis versus a simple alternative hypothesis, it can be observed

that Bayes' factor is just the likelihood ratio of H_p to H_d.

When a parameter θ is continuous (e.g. a proportion or a mean), one or both of the two competing hypotheses may be composite. Consider testing $H_p : \theta \leq \theta_0$ versus $H_d : \theta > \theta_0$ (so, if θ is a proportion, $\Theta_p = [0, \theta_0]$ and $\Theta_d = (\theta_0, 1]$), and let $\pi(\theta)$ denote the prior probability density describing available knowledge about θ. Accordingly, in contrast to the discrete probabilities earlier of $\pi_p = \Pr(\theta = \theta_0)$ and $\pi_d = \Pr(\theta = \theta_1)$, the prior probabilities π_p and π_d are now:

$$\pi_p = \Pr(\theta \in \Theta_p) = \int_{\Theta_p} \pi(\theta) d\theta;$$

$$\pi_d = \Pr(\theta \in \Theta_d) = \int_{\Theta_d} \pi(\theta) d\theta.$$

The posterior probability of hypothesis H_p can be computed as

$$\alpha_p = \Pr(\theta \in \Theta_p \mid x) = \int_{\Theta_p} \pi(\theta \mid x) d\theta$$

$$= \int_{\Theta_p} f(x \mid \theta) \pi(\theta) d\theta / f(x), \tag{7.53}$$

where $f(x) = \int_{\Theta} f(x \mid \theta) \pi(\theta) d\theta$ is the normalising constant. Similarly, the posterior probability of the alternative hypothesis H_d is of the form

$$\alpha_d = \int_{\Theta_d} f(x \mid \theta) \pi(\theta) d\theta / f(x). \tag{7.54}$$

The posterior odds whenever hypotheses are composite is given by

$$\frac{\alpha_p}{\alpha_d} = \frac{\int_{\Theta_p} f(x \mid \theta)\pi(\theta)d\theta}{\int_{\Theta_d} f(x \mid \theta)\pi(\theta)d\theta}. \qquad (7.55)$$

It will now be shown that in the more general case of testing composite hypotheses, the Bayes' factor depends also on the prior input and is not equivalent to a likelihood ratio. To show this, it is useful to rewrite the prior density $\pi(\theta)$ in the following way. Let $\pi_{H_p}(\theta)$ denote the restriction of the prior density on Θ_p, and $\pi_{H_d}(\theta)$ denote the restriction of the prior density on Θ_d, that is,

$$\pi_{H_p}(\theta) = \frac{\pi(\theta)}{\pi_p} \text{ for } \theta \in \Theta_p \ ;$$

$$\pi_{H_d}(\theta) = \frac{\pi(\theta)}{\pi_d} \text{ for } \theta \in \Theta_d.$$

The probability densities $\pi_{H_p}(\theta)$ and $\pi_{H_d}(\theta)$ describe how the prior probability is spread over the competing hypotheses. In other words they are the conditional densities of θ given H_p and H_d, respectively. Therefore, the prior density, $\pi(\theta)$, can be written as

$$\pi(\theta) = \begin{cases} \pi_p \pi_{H_p}(\theta) & \text{if } \theta \in \Theta_p \\ \pi_d \pi_{H_d}(\theta) & \text{if } \theta \in \Theta_d \end{cases}.$$

The posterior probabilities (7.53) and (7.54) are easily rewritten as

$$\alpha_p = \pi_p \int_{\Theta_p} f(x \mid \theta)\pi_{H_p}(\theta)d\theta / f(x)$$

and

$$\alpha_d = \pi_d \int_{\Theta_d} f(x \mid \theta)\pi_{H_d}(\theta)d\theta / f(x).$$

Bayes' factor being the ratio between the posterior odds and the prior odds simplifies to

$$\text{BF} = \frac{\int_{\Theta_p} f(x \mid \theta)\pi_{H_p}(\theta)d\theta}{\int_{\Theta_d} f(x \mid \theta)\pi_{H_d}(\theta)d\theta}.$$

Bayes' factor is now the ratio of weighted likelihoods under the postulated hypotheses, and it no longer depends only on the sample data. The prior input is via the weights $\pi_{H_p}(\theta)$ and $\pi_{H_d}(\theta)$.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

8

Assessment of the Performance of Methods for the Evaluation of Evidence

8.1 INTRODUCTION

It is essential for the evaluation of evidence based on a likelihood ratio that the method used to calculate the likelihood ratio performs well in some sense. Previous chapters have introduced methods for the calculation of likelihood ratios. It is important to know how well a particular method performs. If it performs badly then it should not be used. For example, the method may give a high probability for strong misleading evidence (Section 8.3.1). Several methods for the

assessment of the performance of methods for the evaluation of evidence are introduced in this chapter.

In a comparison problem, the likelihood ratio is a function of the measurements of control and recovered data given (at least) two propositions. The two propositions could be those of same-source and different-source. In a discrimination problem, the likelihood ratio is a function of the measurements of one set of data. The two propositions are those of membership of one population *versus* membership of another population. The likelihood ratio is a function of the measurements of control and recovered data, in a comparison problem, or of measurements of one set of data in a discrimination problem. If the data change, the value changes.

The value of a likelihood ratio used in court is a particular value (Section 2.4.3). If the control and recovered data are from the same source in the comparison scenario, or from the first population in the discrimination scenario, the likelihood ratio is expected to be greater than 1. If the control and recovered data are from different sources in the comparison scenario, or from the second population in the discrimination scenario, the likelihood ratio is expected to be less than 1. In practice, it is not known whether the data come from the same or different sources or from a particular population. A value greater than 1 is support for the proposition of same source or membership of the first population. A value less

than 1 is support for the proposition of different sources or membership of the second population.

Performance measures are not for use in court. Their use is for validation purposes in forensic laboratories and in experimental studies. Their use provides a theoretical criterion to determine whether a given method should or should not be used for the evaluation of evidence. See Ramos-Castro et al. (2013) and Zadora et al. (2014).

From the perspective of forensic statistics, the likelihood ratio is the best approach to the evaluation of evidence (Section 2.4.2). Good (1989) has presented an argument to show that it is the only approach to evaluate evidence. Its use provides a balanced and coherent method for determining the support of the evidence for one proposition over another. Its use enables the prior belief of a fact-finder about the truth or otherwise of a proposition to be updated in the light of new evidence to provide a posterior belief. This posterior belief can, in turn, become a prior belief for more evidence and this belief is then, in turn, updated to form a new posterior belief.

Whilst forensic statisticians consider it the only approach to balanced reporting, there are many different forms a procedure based on the likelihood ratio can take. There are five main components in the formulation of a likelihood ratio, the prosecution proposition H_p, the defence proposition H_d, the evidence E itself, the background information I, and the probabilistic model, ideally informed by data, for assessment of the

probabilities or probability densities in the numerator and denominator. There are choices to be made for each of these components.

For propositions, there are four levels in the hierarchy of propositions, offence, activity, source, and sub-source (Chapters 5 and 6). There are various choices for the defence proposition at each level depending on key issues (5.2). The defence could identify someone else as the criminal, it could suggest the criminal is a member of some sub-population other than the sub-population of the defendant, or it may offer no suggestion as to who the criminal is. At the activity and offence levels there are a series of probabilities to be assigned. Examples of these for the evaluation of trace evidence are probabilities for transfer and persistence at activity level (Section 5.3.2) and for relevance at the offence level (Section 5.3.3).

For background information, the information available to (or permissible for) the forensic scientist to use differs from that available to (or permissible for) the fact-finder to use. It has been shown (Section 2.4.4) that, though these two sets of background information differ, there is no effect on the posterior odds available to the fact-finder.

However, the greatest choice lies with the choice of probabilistic model. For the purposes of illustration, the evidence discussed in this chapter will be in the form of measurements of trace evidence, such as the chemical composition of drugs and the elemental composition of glass for the comparison problem and the quantity of drugs on banknotes

for the discrimination problem. A case study of kinship determination (Section 8.5) is used to illustrate graphically some of the methods of assessment. There is randomness associated with the evidence. As mentioned earlier, there are two types of investigation that may be conducted. The first is one of comparison of source for control and recovered evidence. The second is one of discrimination of one piece of evidence as to which of several populations it might belong.

For the evaluation of a comparison, there are three components to the evidence. First, there is control material whose source is known. Second, there is recovered material whose source is unknown. Third, there are background data, chosen according to some notion of relevance, whose sources are known and which are used as training data for the development of the models from which the likelihood ratio in a particular case can be calculated. The assessment of performance considers the values of the likelihood ratio using members of the background data in a manner described in Section 8.4.

Evaluation is also required in a problem of discrimination (Section 7.7) where it is of interest to evaluate the support of the evidence for the assignment of evidential material to one population or another. For example, one may wish to evaluate the support of the evidence of drugs on banknotes for the assignment of the notes to the population of notes associated with a person associated with crime or to the population of notes associated with

a person not associated with crime. With the problem of discrimination, there are training data that can be used to develop statistical models of population membership for each of the possible populations to which the evidence can be assigned. The assessment of performance considers the value of the likelihood ratio using members of each of the populations.

The methods described in previous chapters for the evaluation of evidence depend crucially on the models chosen for the evaluation. The forensic scientist has to make a choice of model. It is reasonable to assume that the choice will be made based on the comparative quality of the models. The scientist will choose as their preferred model the one that is best according to certain criteria. In order to do this, a definition of quality and a method by which it can be assessed are needed. Quality is defined here as the ability of the model to support the correct result as assessed using the training data. Support is defined as the value of the likelihood ratio. A large value of the likelihood ratio is strong support for a proposition, a small value, close to but greater than 1, is weak support for the same proposition.

Various measures of assessment of quality are described here. Evidence E is evaluated with the likelihood ratio (V), or a function of V. In a comparison problem, the evidence E has two components, control evidence E_c and recovered evidence E_r, so that $E = \{E_c, E_r\}$. For the purpose of assessment, the propositions that are considered

here are source propositions. Assessment is a generic measure of the quality of a model; it is not applicable to a particular case and therefore is not for use in court. Thus it is not possible to consider activity or offence propositions for which considerations of transfer, persistence, recovery, and relevance are important but also specific to an individual case. The source propositions discussed here are H_p, that of a common source for E_c and E_r, normally associated with the prosecution, and H_d, that of a different source for E_c and E_r, normally associated with the defence. Then the value V of the evidence E is given by (2.15)

$$V = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)},$$

where I is background information.

A value of V greater than 1 is said to support H_p. A value of V less than 1 is said to support H_d. For any particular case, it is not known which of these propositions is correct.

Consider source propositions and two sets of trace evidence in a comparison problem. One set (E_c) of trace evidence is from a known source and one set (E_r) from an unknown source. The prosecution proposition H_p is that E_c and E_r are from the same source. The defence proposition H_d is that E_c and E_r are from different sources. Assume the existence of a database, known as a *validation database*, which may be used to check the outcomes of the evaluations. In the validation database sets of measurements of trace evidence

of a hierarchical form (Sections 7.3.2 and 7.6.5), within and between groups, are available. Ideally, the validation database is different from the *training database*. Often, however, the training database and the validation database are the same. An example of such trace evidence is that of window glass evidence, with measurements of refractive index and elemental compositions from fragments within the same windows and with many windows that provide the opportunity to compare measurements between windows. Comparisons may then be made of measurements on two sets of different fragments from within the same window (same-source comparisons) and on two sets of different fragments from different windows (different-source comparisons). For each comparison one set may be chosen as the one of known source (E_c) and the other may be chosen as the one of unknown source (E_r). As the source (window) of each set is known, the correct proposition in any comparison, H_p, same-source, or H_d, different-source, is known. For each comparison and given a model to deal with the hierarchical structure, a likelihood ratio LR using a chosen model is calculated. A value of $\log(V)$ greater than 0 is said to support the prosecution proposition H_p. A value of $\log(V)$ less than 0 is said to support the defence proposition H_d. However, unlike a court case where it is not known which proposition is correct, the correct answer is known. A correct answer is one in

which $\log(V)$ is greater than 0 in a same-source comparison or less than 0 in different-source comparison. Performance of the implemented model can then be assessed with a comparison of the results (supports) with the type of comparison (same-source or different-source), which had been made. It is possible that a same-source comparison may result in a negative $\log(V)$; this result is known as a false negative. Similarly, a different-source comparison may result in a positive $\log(V)$; this result is known as a false positive. Note that this procedure can be considered as part of a pre-assessment procedure to answer questions of the following kind: 'Is it possible to obtain a value supporting the hypotheses of interest in this scenario?' (Taroni et al., 2016).

Quality has been defined as the ability of the model to produce likelihood ratios that support the correct proposition. The ability of the model to support the correct proposition is itself defined with a measure of assessment of the performance of the model. Properties of methods for the evaluation of the likelihood ratio are discussed in Section 8.2. Some general ideas relating to assessment and to presentation are discussed in Section 8.3. Methods for the assessment of the performance of the likelihood ratio are discussed in Section 8.4. These are illustrated graphically with an example on kinship determination taken from Taroni et al. (2005) in Section 8.5.

8.2 PROPERTIES OF METHODS FOR EVALUATION

Four distinct properties are mentioned regularly in the context of the role of scientific evidence in court. These are the properties of accuracy, precision, reliability, and validity. It is of interest to consider their definitions in the *Oxford English Dictionary*[1] in the context of evidence evaluation and the likelihood ratio.

- *Reliability*: Ability to be relied on with confidence; trustworthiness, sureness, reliability; to *rely* is to depend on with full trust or confidence.
- *Accuracy*: The closeness of a measurement, calculation, or specification to the correct value. Contrasted with precision (the degree of refinement of the measurement, etc.).
- *Validity*: The quality of being well-founded on fact, or established on sound principles, and thoroughly applicable to the case or circumstances; soundness and strength (of argument, proof, authority, etc.).
- *Precision*: The degree of refinement in a measurement, calculation, or specification, especially as represented by the number of digits given. Contrasted with accuracy (the closeness of the measurement, etc., to the correct value).

For the likelihood ratio method of evidence evaluation, these properties may be interpreted as

[1] http://www.oed.com.

- *Reliability*: The method has the characteristics of trustworthiness and sureness. In 1993, the US Supreme Court ruled that scientific knowledge will assist the trier of fact only if it is also *reliable*, or trustworthy.

> *The requirement that an expert's testimony pertain to 'scientific knowledge' establishes a standard of ... evidentiary reliability - that is, trustworthiness.*
>
> *In a case involving scientific evidence,* evidentiary reliability *will be based on* scientific validity. *(*Daubert v. Merrell Dow Pharmaceuticals.*)*

Note that reliability is interpreted by Royall (2000) to be the probability of observing strong misleading evidence (Section 8.3.1). This is a technical mathematical definition and is not to be confused with the legal terminology.

Note also the following in discussion of an experimental study of complex DNA profiles

> *The findings of this study further demonstrate ... that regardless of the strength or complexity of the DNA data, as long as the models used to analyse the data are reliable, then the LR produced will also be reliable. (Taylor et al., 2015, p. 170)*

Taylor (2014) gives a fuller discussion of the meaning of reliability in DNA profiling.

- *Accuracy*: It is not possible to determine accuracy as there is no true value of the likelihood ratio against which accuracy can be determined. Ramos-Castro and Gonzalez-Rodriguez (2013) define accuracy as the closeness of a

validation ECE curve to the PAV ECE curve (Section 8.5).

- *Validity*: The method is well-founded on fact, is established on sound principles, and is thoroughly applicable to the case or circumstances; it is sound and has strength of argument, proof, and authority. For example, methods relying on probabilistic arguments need to satisfy the laws of probability.
- *Precision*: The method can produce as many digits as is desired though it is restricted to the precision of any measuring instrument used to provide the relevant data; the language is exact – the evidence is so many times more probable if one proposition is true than if the other proposition is true. Precision is also a statistical term defined as the reciprocal of the variance of a distribution. Further discussion of the meaning of precision is given in Biedermann et al. (2016b), which also discusses the concept of *resolution*:

> *'There may be discussion about whether it is feasible and desirable to determine if a single probability should be given to, for example, four significant figures and hence make a distinction between 0.0101 and 0.0102. This is a question of the level of resolution, not of precision. The probabilities 0.0101 and 0.0102 are different and hence, by definition, express different levels of uncertainty. However, on practical grounds, evidential value should be considered as measurable in increments which are no*

> *smaller than that which may be thought discernible by a juror or other fact-finder. This connects to the historical notion of ban, used to name units of information, with deciban being considered as the smallest change in evidential value that is perceptible to humans (Hodges, 1992). Consider a probability value of 1 as a numerator for a likelihood ratio. For denominators of 0.01, 0.0101, and 0.0102, the evidential values are 100, 99.0099, and 98.03922, respectively. It is questionable whether a human being could separate meaningfully these three different evidential values' (p. 394).*

Thus, methods of evaluation of evidence based on the likelihood ratio are reliable and valid. It is not possible to assess accuracy. Precision is subject to the precision of any measuring instrument.

The assessment of these characteristics is not so clear when consideration is of the numbers calculated in a particular case. The assessment has to take account of the relevance of the data and model used to produce the numbers. Taylor et al. (2016c) stated that

> *'Obviously, the more structured data we have, the better. But, in real life, the numbers of experiments that can be carried out are limited. It is thus important to know if/when our knowledge is sufficient and when one needs to perform further experiments to be in a position to report the value of the observations made' (p. 402).*

The method may be reliable and valid. The data have to be assessed separately for reliability and validity.

The meanings of these words are also discussed in Robertson et al. (2016, p. 62) along with sensitivity and specificity. In the context of evidence evaluation, sensitivity is the probability of a likelihood ratio greater than 1 when the proposition in the numerator is true. Specificity is the probability of a likelihood ratio less than 1 when the proposition in the denominator is true. The existence of considerable discussion indicates that these words should be used with care.

Evidence in a particular case is evaluated by the likelihood ratio. This is reported as a single value for reasons explained in Section 2.4.3. Further precision in the sense of closeness of several values is not an issue since only one value has been calculated. The calculation, especially if it involves a computer program, may return a result to many significant figures. However, those beyond the precision of the measuring device that provided the original data will be meaningless. The precision of a model may be determined. The precision of a particular result provided by the model may not be determined.

Whilst it is not possible to consider properties of the value of the evidence in a particular case, it is possible to assess the properties of the model that was used to determine the value.

For investigation of a particular case, there may be a choice of models for the evaluation of evidence. It will strengthen the expert's report

if they can testify that they have used the best model available. In order to be able to so testify, the performances of various competing methods have to have been assessed.

8.3 GENERAL TOPICS RELATING TO SAMPLE SIZE ESTIMATION AND TO ASSESSMENT

Two general topics are discussed. The first topic concerns the choice of an optimal sample size in a comparison problem for continuous measurements through determination of the *probability of strong misleading evidence*. The second topic concerns a method for assessment of performance known as *calibration*.

8.3.1 Probability of Strong Misleading Evidence: A Sample Size Problem

This section discusses probabilistic limits and criteria for the sample size when collecting evidence in the form of continuous measurements. Consider the following quote:

> *Statistical thinking concerns the relation of quantitative data to a real-world problem, often in the presence of variability and uncertainty. It attempts to make* precise and explicit *(current authors' italics) what the data has (*sic*) to say about the problem of interest. (Mallows, 1998, p. 3)*

Two of the requirements of the courts in the United States for scientific evidence are that it be relevant and reliable. Likelihood ratios consider the matter of relevance. Evidence can be said to be relevant if the likelihood ratio is different from 1; i.e. the posterior odds after presentation of the evidence is different (larger or smaller) from the prior odds before presentation of the evidence. In a book and a series of papers Mellen and Royall (Royall 1997, 2000; Mellen 2000; Mellen and Royall 1997) discuss the issue of reliability through the concept of weak evidence and strong misleading evidence. Weak evidence is evidence with a low likelihood ratio. Strong misleading evidence is evidence with a high likelihood ratio in favour of the wrong proposition; e.g. evidence which has a high likelihood ratio in favour of the prosecution proposition when the defence proposition is true.

Relevance is addressed in the United States through Rule 401 of the Federal Rules of Evidence. Evidence is relevant if 'it has any tendency to make a fact more or less probable than it would be without the evidence; and the fact is of consequence in determining the action'.

A change in the odds, greater or less than previously, in favour of the prosecution's proposition, through a value for the evidence different from 1, is a change in the probability of the prosecution's proposition. Thus, there is a connection

between Rule 401 and the likelihood ratio. Note that this concept of relevance is different from that discussed in Chapters 5 and 6 where it is taken to be the probability that trace evidence, which is recovered from the victim/PoI and which matches (in some sense) trace evidence associated with the PoI/victim is connected with the crime (Stoney, 1991a,1994), (Evett et al., 1998). The relevance that is a probability is used as a term in the expression for the likelihood ratio. The relevance as defined in Rule 401 of the Federal Rules of Evidence can be thought of in terms of the value of the likelihood ratio as a whole.

Rule 702 of the Federal Rules of Evidence lays out when expert witnesses may be allowed to testify. 'A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- the expert's scientific, technical, or other specialised knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- the testimony is based on sufficient facts or data;
- the testimony is the product of reliable principles and methods; and
- the expert has reliably applied the principles and methods to the facts of the case'.

In 1999, the US Supreme Court stated

> Daubert's *general principles apply to the expert matters described in Rule 702. The Rule, in respect to all such matters 'establishes a standard of evidentiary reliability' (509 U.S. at p. 590)*
>
> *. . . the trial judge must determine whether the testimony has 'a reliable basis in the knowledge and experience of [the relevant] discipline' (509 U.S. at p. 592). (*Kumho Tire Co., Ltd. v. Carmichael*)*

Consider two competing propositions, A and B for evidence E. The likelihood ratio is $\Pr(E \mid A)/\Pr(E \mid B)$. Denote this as V_{AB}. Strong evidence in favour of A will be taken to be evidence for which the likelihood ratio is greater than a specified value $k(> 1)$, say. Thus the probability of strong misleading evidence will be evidence for which $\Pr(V_{AB} > k)$ when B is the correct proposition or $\Pr(V_{AB} < 1/k)$ when A is the correct proposition. The subscript B can be used in the notation to clarify under which proposition the probability is being determined, so that the probability of strong misleading evidence when B is the correct proposition can be denoted $\Pr_B(V_{AB} > k)$. Consider E as a set of measurements x (with corresponding random variable X), such as DNA profiles (PoI, victim, and background data) or refractive indexes of glass (PoI, crime scene, and background data). Then, it can be shown (Royall, 1997) that it is unlikely there will be strong evidence favouring A,

when B is true. In particular,

$$\begin{aligned}&\Pr_B(V_{AB} > k)\\&\quad = \Pr_B(\Pr(X = x \mid A)/\Pr(X = x \mid B) > k) < 1/k,\end{aligned}$$

where $X = x$ has been substituted for E.

Consider the set S of all possible values of X, which produce a value V_{AB} greater than k. For each of these values x of X,

$$\Pr(X = x \mid B) < \Pr(X = x \mid A)/k$$

by rearrangement of the probabilistic inequality earlier. It is then possible to sum over all values x of X in S to obtain

$$\Pr(S) = \sum_{x \in S} \Pr(X = x \mid B) < \sum_{x \in S} \Pr(X = x \mid A)/k.$$

The right-hand sum $\sum_{x \in S} \Pr(X = x \mid A)$ will not be greater than 1, as it is a sum of mutually exclusive probabilities. Thus, $\sum_{x \in S} \Pr(X = x \mid A)/k < 1/k$ and, hence $\Pr(S) < 1/k$. Further details and a stronger result that 'if an unscrupulous researcher sets out deliberately to find evidence supporting his favourite but erroneous hypothesis over his rival's, which happens to be correct, by a factor of k, then the chances are good that he will be eternally frustrated' are given in Royall (1997).

Values of k of 8 and 32 are proposed by Royall (1997, p. 25) to represent 'fairly strong' and 'strong' evidence respectively. These are justified with reference to an urn example (see Section 1.7.2). Consider an urn that may contain all white balls or half white balls and half black balls. If three balls are drawn without replacement and all are found to be white this may be thought of as 'fairly strong' evidence that the urn contains only white balls. The probability of this event if the urn contains half white and half black balls is $(1/2)^3 = 1/8$. If five balls are drawn without replacement and all are found to be white, this may be thought of as 'strong' evidence that the urn contains only white balls. The probability of this event if the urn contains half white and half black balls is $(1/2)^5 = 1/32$. Similar benchmarks have been proposed by Edwards (1992), Jeffreys (1983), and Kass and Raftery (1995). Comparison should also be made with Sections 2.4.6 and 4.2.1.

An example is given in Mellen (2000) of the application of these ideas to DNA evidence. Let s denote the source of the DNA and d denote the defendant. Consider two propositions

- The defendant is the source of the crime scene DNA ($s = d$);
- An unknown person is the source ($s \neq d$).

Suppose that the crime scene DNA and the DNA of the PoI have corresponding features. Let z denote the genotype that is observed. Let Z_i denote the random variable corresponding to

the genotype from person i; Z_r is the genotype from the source (recovered) of the DNA at the crime scene and Z_c is the genotype from the PoI (control). Then $Z_c = Z_r = z$. The probability of misleading evidence (evidence whose value V is greater than k) is evaluated assuming $s \neq d$. Thus

$$\begin{aligned}
&\Pr(V > k \mid Z_r = z) \\
&\quad = \Pr(V > k, Z_c = z \mid Z_r = z) \\
&\quad = \Pr(V > k \mid Z_c = z, Z_r = z) \Pr(Z_c = z \mid Z_r = z) \\
&\quad < \Pr(Z_c = z \mid Z_r = z).
\end{aligned}$$

Assuming that $s \neq d$, this final probability is equal to the conditional genotype probability. The probability of strong misleading evidence is not greater than the conditional genotype probability. As stated in Mellen (2000), 'as might be expected, if the genotype z tends to be rare among individuals in the same genetic subset of the population as the defendant, then the probability of observing genotypes in the defendant and the reference sample that constitute strong misleading evidence is not great. If, on the other hand, the genotype z tends to be quite common in this subpopulation, then the probability might be larger'. (p. 140)

In the discussion at the end of Mellen and Royall (1997) comment is made on several useful features of the analysis. These include the following:

- *Separation* between measures of evidence and reliability of the process that produces the evidence;

- *Distinction* between the strength of implicating evidence and the improbability of its occurrence – there is a low probability of misleading strong implicating evidence;
- *Explicit conditioning* on the circumstances of a case: condition on the non-DNA evidence to delimit the suspect population and condition on the source DNA type in probabilities of strong implicating evidence;
- *Generality* of the methods, the importance of conditional probabilities (Balding and Donnelly, 1995a), and the extension of the methods to identification evidence other than DNA evidence.

Royall (2000) extended these ideas to continuous data and discrimination. Consider two propositions H_1 and H_2 for evidence in the form of measurements X, such that, for H_1, denoted f_1,

$$X \sim N(\theta_1, \sigma^2)$$

and, second, for H_2, denoted f_2,

$$X \sim N(\theta_2, \sigma^2).$$

Let there be data $x_1, \ldots, x_n$. Then the likelihood functions f_{1n} and f_{2n}, in the two propositions, are

$$f_{2n} = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \theta_2)^2\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_2)^2 \right\}$$

$$f_{1n} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_1)^2 \right\}$$

$$\frac{f_{2n}}{f_{1n}} = \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \theta_2)^2 - \sum_{i=1}^{n} (x_i - \theta_1)^2 \right] \right\}$$

$$= \exp\left\{ \frac{n(\theta_2 - \theta_1)}{\sigma^2} \left(\bar{x} - \frac{\theta_1 + \theta_2}{2} \right) \right\}.$$

If H_1 is true then

$$\bar{X} \sim N(\theta_1, \sigma^2/n)$$

and it can be shown (Royall, 2000) that

$$\Pr_1 \left(\frac{f_{2n}}{f_{1n}} > k \right) = \Phi\left(-\frac{\Delta\sqrt{n}}{2\sigma} - \frac{\sigma \log_e(k)}{\Delta\sqrt{n}} \right)$$

where $\Delta = \mid \theta_2 - \theta_1 \mid$ and the subscript 1 associated with the Pr indicates that the first proposition is taken to be true. In analogous notation, $\Pr_2$ will indicate that the probability is to be determined assuming the second proposition to be true. If Δ expressed as a multiple c of the standard error of $\bar{X}$ is such that $\Delta = \mid \theta_2 - \theta_1 \mid = c\sigma/\sqrt{n}$, then

$$\Pr_1 \left(\frac{f_{2n}}{f_{1n}} > k \right) = \Phi\left(-\frac{c}{2} - \frac{\log_e(k)}{c} \right),$$

assuming θ_1 to be the true mean. This function is a so-called *bump* function.

If θ_1 is the true mean then there is very little chance of observing strong evidence supporting θ_2 over θ_1 when the difference Δ between the two parameter values is a small fraction of the standard error $\sigma/\sqrt{n}$.

These ideas may be used to determine a sample size based on the criteria of controlling for the probability of strong misleading evidence and for the probability of weak evidence. Consider a likelihood ratio f_2/f_1 of density functions where the subscripts denote the two competing propositions.

- *Strong* evidence is defined as evidence for which the f_2/f_1 is greater than a pre-specified value k, or, conversely, less than $1/k$.
- *Strong misleading* evidence is defined as evidence for which f_2/f_1 is greater than the pre-specified value k when proposition 1 is assumed true, or, conversely, less than $1/k$ when proposition 2 is assumed true.
- *Weak* evidence is evidence that is not strong; i.e. evidence for which $1/k < f_2/f_1 < k$.

The probability, $M(n)$, of observing strong misleading evidence, as a function of the sample size n, is given by

$$\begin{aligned} M(n) &= \Pr_1\left(\frac{f_{2n}}{f_{1n}} > k\right) \\ &= \Pr_2\left(\frac{f_{1n}}{f_{2n}} > k\right) \\ &= \Phi\left(-\frac{\Delta\sqrt{n}}{2\sigma} - \frac{\sigma\log_e(k)}{\Delta\sqrt{n}}\right) \end{aligned} \qquad (8.1)$$

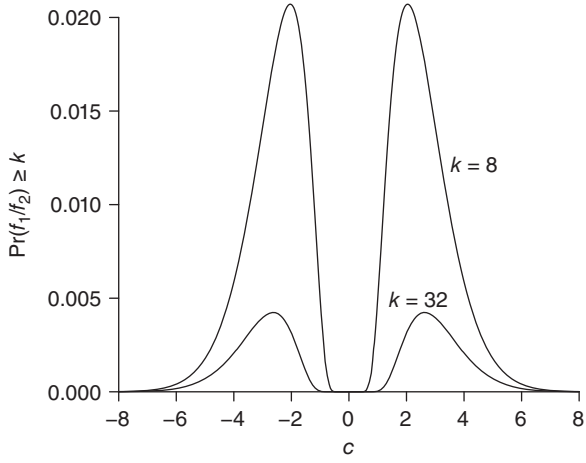

Figure 8.1 Bump function for the probability of misleading evidence $\Pr_1(\frac{f_{2n}}{f_{1n}} > k)$ for $k = 8$ and $k = 32$ as a function of c, the distance from the true mean to the alternative, in standard errors. Source: Royall (2000). Reprinted with permission of Taylor and Francis Ltd.

and the probability $W(n)$, of observing weak evidence, as a function of the sample size n, is given by

$$\begin{aligned} W(n) &= \Pr_1\left(1/k < \frac{f_{2n}}{f_{1n}} < k\right) \\ &= \Pr_2\left(1/k < \frac{f_{2n}}{f_{1n}} < k\right) \\ &= \Phi\left(-\frac{\Delta\sqrt{n}}{2\sigma} + \frac{\sigma\log_e(k)}{\Delta\sqrt{n}}\right) - M(n). \quad (8.2) \end{aligned}$$

Consider an example where the characteristic of interest is the refractive index of glass. A window is broken at the crime scene. A PoI is apprehended soon afterwards and they have fragments of glass

on their clothing. They explain the presence of the fragments by saying that they had just broken a glass. The two propositions of interest are

H_1: The glass fragments on the clothing of the PoI came from the crime scene window.

H_2: The glass fragments on the clothing of the PoI came from the broken glass.

The window at the crime scene is of a very common type. There is a population database with a refractive index of assumed known mean (θ_1) and standard deviation σ for variation within windows.[2] The glass from which the broken glass was manufactured is also of a common type with a refractive index of assumed known mean (θ_2) and standard deviation σ for variation within glasses (the same as for the crime scene window glass).

A pre-assessment question (see Section 5.5.2 for further discussion) is the determination of the number of fragments of glass from the clothing of the PoI to be examined. Once this number has been determined, refractive indices of fragments of glass from the clothing of the PoI and fragments from the crime scene window can be measured and compared using the likelihood ratio expressions of Chapter 7.

From (8.1) and (8.2), with $\Delta = \mid \theta_1 - \theta_2 \mid$ and σ known, the two unknowns are the sample size n and the criterion k for strong evidence. Both

[2]This scenario is different from the hierarchical model used for evidence evaluation in Section 7.3.

n and k can be varied and the corresponding values of $M(n)$ and $W(n)$ investigated. For the determination of the sample size in the pre-assessment stage it is necessary to consider three criteria:

- the meaning of 'strong' (the value for k), and, as a consequence:
- the probability of strong misleading evidence;
- the probability of weak evidence.

This procedure is illustrated using the following values for the parameters:

- θ_1: 1.5195073,
- θ_2: 1.5195730,
- σ: 0.0000492.

Then, $\Delta = 0.0000657$ and from (8.1) and (8.2), Table 8.1 can be obtained.

Suppose it is decided that strong evidence (either in support of H_1 or of H_2) is evidence with a value greater than 8, and that it is tolerable to have a probability of strong misleading evidence no greater than 0.005 and to have a probability

Table 8.1 Probabilities of strong misleading evidence $M(n)$ and weak evidence $W(n)$ for boundary values k of 8 and 32 for strong evidence and sample sizes n of 5, 10, and 20

k	$M(5)$	$W(5)$	$M(10)$	$W(10)$	$M(20)$	$W(20)$
8	0.0143	0.1985	0.0046	0.0481	0.0004	0.0038
32	0.0040	0.3658	0.0017	0.0967	0.0002	0.0079

of weak evidence no greater than 0.05. These criteria will be satisfied with a sample of size 10. This follows from inspection of Table 8.1, in the row for $k = 8$ and the columns for $M(10)$ and $W(10)$ where the corresponding cell values are 0.0046 (<0.005) for $M(10)$ and 0.0481 (<0.05) for $W(10)$.

A probability of 0.005 for strong misleading evidence is the probability that strong evidence will be obtained that the glass fragments on the clothing of the PoI came from the crime scene window, when in fact they came from the broken glass or that the glass fragments on the clothing of the PoI came from the broken glass when in fact they came from the crime scene window.

Other applications can be considered. For example, consider the sampling of a consignment of drugs as described in Sections 4.3 and 4.4. The sample size n is determined by the criterion of satisfying a pre-specified probability that the true proportion of illicit drugs in the consignment is greater than a pre-specified value.

In contrast to that criterion, consider two propositions about the possible source of the consignment:

H_1: The drugs are from a source with mean quantity of drug per tablet of θ_1;

H_2: The drugs are from a source with mean quantity of drug per tablet of θ_2.

The criterion k for strong evidence can be chosen. Then the probabilities for strong misleading

evidence and for weak evidence may be determined.

A procedure for estimating the quantity of drug in a consignment, based on a sample from the consignment, was described in Section 4.6. This procedure could be adapted to estimate the mean amount of drug in one tablet in the consignment.

There are, thus, two procedures. In the first, probabilities for strong misleading evidence and for weak evidence are used to determine sample size in a pre-assessment stage before sampling to compare two propositions about the source of the consignment. In the second, an estimate of the quantity of drug in the consignment is obtained, without reference to any possible source of the consignment.

The results from the sampling can be used in the determination of a likelihood ratio to evaluate the evidence from the consignment in support of either H_p or H_d.

It is also possible to determine a sample size from the criteria based on (8.1) and (8.2). This may give a different sample size from that obtained using the criteria described in Section 4.3. However, the two criteria are designed to answer two different questions and thus may give different answers. The criteria based on (8.1) and (8.2) are designed to compare two propositions about a mean value. The criterion described in Section 4.3 is designed to satisfy a pre-specified probability that the true proportion of illicit drugs in the consignment is greater than a pre-specified value.

The ideas discussed here are for consideration of model performance. They are not for application in a particular case. Thus it makes sense to discuss the probability of a likelihood ratio which it would not make sense to do in the context of a particular case (see Sections 2.4.3 and 7.6.3).

8.3.2 Calibration

The performance of a procedure can be characterised by the distribution of likelihood ratios, one for each of two propositions. An interesting property of the likelihood ratio is that the likelihood ratio of the likelihood ratio is itself the likelihood ratio (van Leeuwen and Brümmer, 2013). Equation (2) of van Es et al. (2017) states

$$\frac{\Pr(\mathrm{LR} \mid H_p)}{\Pr(\mathrm{LR} \mid H_d)} = \mathrm{LR}.$$

The proof is in van Leeuwen and Brümmer (2013) and their notation and example is used here. A speaker recognition system has as input two speech segments, denote these as X and Y. Let $s = f(X, Y)$ be a single, scalar score (see Section 7.8). The likelihood ratio, here denoted r for consistency with van Leeuwen and Brümmer (2013), is a function of s:

$$r = \frac{\Pr(s \mid H_p, M)}{\Pr(s \mid H_d, M)}, \tag{8.3}$$

where H_p is the proposition that X and Y originate from the same speaker, H_d is the proposition that

X and Y are from different speakers, and M is a probabilistic model for s. In practice s is a scalar score, though the authors claim the theory is sufficiently general to remain applicable for more ambitious models.

Let $\Pr(H_p \mid M) = \pi$. Then

$$\begin{aligned}\Pr(H_p \mid s, M, \pi) &= \frac{\Pr(s \mid H_p, M, \pi)\pi}{\Pr(s \mid H_p, M, \pi)\pi + \Pr(s \mid H_d, M, \pi)(1-\pi)} \\ &= \frac{r\pi}{r\pi + (1-\pi)}, \end{aligned} \tag{8.4}$$

since $\Pr(s \mid H_p, M, \pi) = \Pr(s \mid H_p, M)$ and $\Pr(s \mid H_d, M, \pi) = \Pr(s \mid H_d, M)$. Thus r is such that the posterior for H_p depends on s only through r. A similar argument holds for H_d with

$$\Pr(H_d \mid s, M, \pi) = \frac{(1-\pi)}{r\pi + (1-\pi)}.$$

Thus, the posterior probability may be written as

$$\Pr(h \mid s, M, \pi) = \Pr(h \mid r, M', \pi), \quad h \in \{H_p, H_d\} \tag{8.5}$$

where M' has been introduced to denote M augmented with (8.3). Then

$$\begin{aligned}\frac{\Pr(H_p \mid s, M, \pi)}{\Pr(H_d \mid s, M, \pi)} &= \frac{\pi}{(1-\pi)} \frac{\Pr(s \mid H_p, M)}{\Pr(s \mid H_d, M)} \\ &= \frac{\pi}{(1-\pi)} r; \\ \frac{\Pr(H_p \mid r, M', \pi)}{\Pr(H_d \mid r, M', \pi)} &= \frac{\pi}{(1-\pi)} \frac{\Pr(r \mid H_p, M')}{\Pr(r \mid H_d, M')}\end{aligned}$$

$$\Rightarrow \frac{\pi}{(1-\pi)} r = \frac{\Pr(H_p \mid s, M, \pi)}{\Pr(H_d \mid s, M, \pi)}$$
$$= \frac{\Pr(H_p \mid r, M', \pi)}{\Pr(H_d \mid r, M', \pi)} \text{ from (8.5)}$$
$$= \frac{\pi}{(1-\pi)} \frac{\Pr(r \mid H_p, M')}{\Pr(r \mid H_d, M')}$$
$$\Rightarrow r = \frac{\Pr(r \mid H_p, M')}{\Pr(r \mid H_d, M')}.$$

As discussed in Section 8.1, the use of a training set enables the determination of distributions for likelihood ratios. Van Es et al. (2017) used probability density estimates from a training set for the distribution of likelihood ratios based on same-source propositions and the distribution of likelihood ratios based on different-source propositions. They used these distributions to amend, or calibrate, the result, s_0, say, in a particular case.

Let r_0 be the likelihood ratio derived from s_0, which is to be calibrated based on the information of probability densities (same-source and different-source), denote these as $f(r \mid H_p)$ and $f(r \mid H_d)$. Then the calibrated likelihood ratio is given by

$$\frac{f(r_0 \mid H_p)}{f(r_0 \mid H_d)}.$$

It is this ratio that van Es et al. (2017) reported as the value of the evidence. A problem with this approach is that it is not possible to update evidential value when new evidence is obtained.

The posterior odds for one piece of evidence become the prior odds for the next piece of evidence. Given two pieces of evidence, E_1 and E_2, from (2.16)

$$\frac{\Pr(H_p \mid E_1, E_2)}{\Pr(H_d \mid E_1, E_2)} = \frac{\Pr(E_2 \mid H_p, E_1)}{\Pr(E_2 \mid H_d, E_1)} \times \frac{\Pr(E_1 \mid H_p)}{\Pr(E_1 \mid H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}. \quad (8.6)$$

Calibration of these likelihood ratios gives two new likelihood ratios, LR_1 and LR_2 defined as

$$\mathrm{LR}_1 = \frac{\Pr\{\Pr(E_1 \mid H_p)/\Pr(E_1 \mid H_d) \mid H_p\}}{\Pr\{\Pr(E_1 \mid H_p)/\Pr(E_1 \mid H_d) \mid H_d\}}$$

and

$$\mathrm{LR}_2 = \frac{\Pr\{\Pr(E_2 \mid E_1, H_p)/\Pr(E_2 \mid E_1, H_d) \mid H_p\}}{\Pr\{\Pr(E_2 \mid E_1, H_p)/\Pr(E_2 \mid E_1, H_d) \mid H_d\}}.$$

It is not clear how to combine evidential value for two or more pieces of evidence that have been calibrated. For example, the product of LR_1 and LR_2 has no meaning. Mathematically it is not possible to model the dependencies of a likelihood ratio for two or more pieces of evidence if one of the likelihood ratio expressions has been adjusted during the calibration process.

The result that the likelihood ratio of the likelihood ratio is the likelihood ratio is a general result. Consider a class of problems such as that of kinship (Section 8.5). The likelihood ratio can

be calculated for all possible outcomes and a distribution obtained. It then makes sense to refer to the probability of a likelihood ratio (or the value of the probability density function). In practice, interest is concentrated on one particular value of the likelihood ratio. In that context, it makes no sense to talk of the probability of the likelihood ratio. The result is an interesting theoretical result but not one that should be used in practice. The result can be used for comparison of performance. For example, plot the likelihood ratio LR against $\Pr(\text{LR} \mid H_p)/\Pr(\text{LR} \mid H_d)$. The closer the plot is to a straight line, the better the method.

This definition of calibration can be contrasted with that used by Bayesian statisticians. For example, DeGroot and Fienberg (1983) discuss calibration with regard to weather forecasts. A forecaster is well-calibrated if their forecast of a probability p of rain the following day is such that on $100p\%$ occasions on which they make that forecast it does indeed rain on the following day.

8.4 ASSESSMENT OF PERFORMANCE OF A PROCEDURE FOR THE CALCULATION OF THE LIKELIHOOD RATIO

Sensitivity analyses of models provide one measure of reliability. The change in the output (LR) of

a model as parameters change can be investigated. See Taylor et al. (2014) for an example with DNA profiles where variation of the value of the LR is investigated as there are changes in the value of F_{ST}, the weights given to genotype combinations in a continuous interpretation model, and the composition of the relevant population. This section does not concern itself with these analyses of sensitivity. Instead it concentrates on the assessment of the performance of a particular model where the characteristics of the model remain unchanged. The changes considered are those of the evidence, the control and recovered data, for example.

Consider a training database where the source of each data point is known (in the comparison scenario) or where the population membership is known (in the discrimination scenario). It is then possible to study the variation of the likelihood ratio as the control and recovered data change.

Consider hierarchical data with two levels in a comparison problem. An example is that of a training set of windows with measurements of refractive indices. There is variation of refractive indices within windows and between windows. (See Section 7.3.2 for an example.) The training set consists of m windows with n fragments of glass in each window. Denote the measurements of refractive indices of the fragments by $x_{ij}, i = 1, \ldots, m, j = 1, \ldots, n$. Two fragments, with measurements denoted y_1 and y_2, represent control and recovered data, respectively. They are chosen

from the training set and evaluated against two propositions, H_p that the two fragments come from the same source and H_d that the two fragments come from different sources and a likelihood ratio $f(y_1, y_2 \mid H_p)/f(y_1, y_2 \mid H_d)$ is calculated. The likelihood ratio for all possible pairs of fragments in the training set can be calculated. A set of within-window likelihood ratios is calculated, consisting of $\frac{1}{2}mn(n-1)$ values (the pairs of fragments compared consist of two fragments from the same window) and a set of between-window likelihood ratios is calculated, consisting of $\frac{1}{2}m(m-1)n^2$ values (the pairs of fragments compared consist of two fragments from different windows). It is to be expected the calculated within-window likelihood ratios are greater than 1 and the between-window likelihood ratios are less than 1. These two sets of likelihood ratios provide the distributions of likelihood ratios for within-source and between-source comparisons. Examples of such distributions in a problem of kinship determination are shown in Figure 8.4. Performance measures are obtained through study of these distributions. These distributions are used to assess the performance of the likelihood ratio as a method for the evaluation of evidence. They are not used in a particular case.

Consider a set $\{E_i = (E_{ci}, E_{ri}), i = 1, \ldots, n_p\}$ of evidential pairs (control c and recovered r) for same-source comparisons and a set $\{E_j = (E_{cj}, E_{rj}),$

$j = 1, \ldots, n_d\}$ of evidence pairs (control and recovered) for different-source comparisons. Log likelihood ratios for same-source comparisons

$$V_{\text{same}} = \log\left\{\frac{f(E_{ci}, E_{ri} \mid H_p}{f(E_{ci}, E_{ri} \mid H_d}\right\}, \quad i = 1, \ldots, n_p$$

and for different source comparisons

$$V_{\text{diff}} = \log\left\{\frac{f(E_{cj}, E_{rj} \mid H_p}{f(E_{cj}, E_{rj} \mid H_d}\right\}, \quad j = 1, \ldots, n_d$$

may be calculated using the ideas of Section 8.1.

A similar set of calculations can be made for the discrimination problem. Let H_A and H_B be the propositions for membership of population A and B, respectively. Consider a set $\{E_{Ai}, i = 1, \ldots, n_A\}$ of pieces of evidence from population A and a set $\{E_{Bj}, j = 1, \ldots, n_B\}$ of pieces of evidence from population B. Log likelihood ratios for members of the set from population A

$$V_A = \log\left\{\frac{f(E_{Ai} \mid H_A}{f(E_{Ai} \mid H_B}\right\}, \quad i = 1, \ldots, n_A$$

and for members of the set from population B

$$V_B = \log\left\{\frac{f(E_{Bj} \mid H_A}{f(E_{Bj} \mid H_B}\right\}, \quad j = 1, \ldots, n_B$$

may be calculated using the ideas of Section 8.1.

8.4.1 Histograms and Tippett Plots

Denote the two propositions to be considered as H_1 and H_2 where $\{H_1, H_2\} = \{H_p, H_d\}$ in the comparison scenario and $\{H_A, H_B\}$ in the discrimination scenario.

Two histograms are determined as estimates of probability distributions of results for log(LR) for data from H_1 and, separately, from H_2. The base of the logarithms is conventionally chosen to be 10 but it need not be. Also, likelihood ratios could be used without a logarithmic scale but this could lead to difficulties of presentation. The discriminating power[3] of a method at a particular value of log(LR) is the amount of overlap of the distributions at that value. If there is no overlap then there is 100% discrimination; this is a practical situation since for continuous data there will always be some overlap. If there is no separation then one distribution is wholly included in the range of the other and there is no discrimination. Examples from a case study of kinship determination are shown in Figure 8.4.

If the training set is not large, then the estimates of the probability distributions can be replaced by histograms. Histograms of $\log_{10}$(LR) values in an artificial set of LR values are shown in Figure 8.2. Histograms from a case study in kinship determination are shown in Figures 8.5 and 8.6.

[3]This definition of discriminating power is not to be confused with that of Section 3.5.

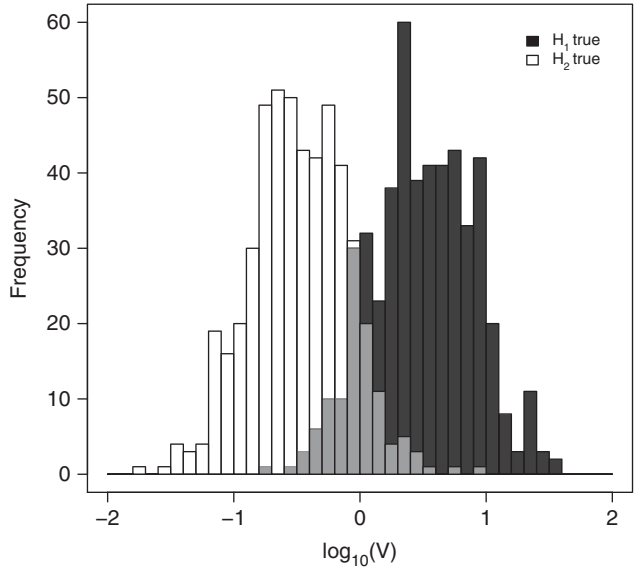

Figure 8.2 Histograms of $\log_{10}$(LR) values in an artificial set of LR values. The area of overlap is indicated in grey.

Tippett plots are generalisations of rates of misleading evidence for comparisons. They are the complement of empirical cumulative distribution functions for same-source and different-source comparisons or for the two populations in a discrimination problem. The plots come in pairs, one for same-source comparisons and one for different-source comparisons or one for membership of group A and one for membership of group B. The log(LR) is plotted on the x-axis and, for a particular value x_0 of the log(LR), the value on the y-axis is the proportion of comparisons greater than x_0. For same-source comparisons, it is to

be hoped that all log(LR) values are greater than 0. Thus for $x < 0$, it is hoped the corresponding value on the y-axis will be 1 (or 100%). Similarly, for different-source comparisons, it is to be hoped that all log(LR) values are less than 0. Thus for $x > 0$, it is hoped the corresponding value on the y-axis will be 0 (or 0%). An artificial example from Zadora et al. (2014) is shown in Figure 8.3. Tippet plots from the case study in kinship determination are shown in Figure 8.8. Other examples of Tippett plots are presented in Meuwly (2001), Riva (2011), Riva and Champod (2014), and Lucena-Molina et al. (2015a).

The distance from the intersection of the same-source plot with the line $\log(LR) = 0$ and the line $y = 1(100\%)$ is the rate of misleading evidence for same-source comparisons, the proportion of same-source comparisons that have a value of $\log(LR) < 0$ ($LR < 1$) (false negatives). The distance from the intersection of the different-source plot with the line $\log(LR) = 0$ and the line $y = 0(0\%)$ is the rate of misleading evidence for different-source comparisons, the proportion of different-source comparisons that have a value of $\log(LR) > 0$ ($LR > 1$) (false positives).

Applications of this procedure have been described for DNA (Evett et al., 1993; Evett and Buckleton, 1996; Evett and Weir, 1998), for speaker recognition (Meuwly and Drygajlo, 2001), for relationships between heroin seizures (Dujourdy et al., 2003), and for marks left on gun cartridges (Riva, 2011; Riva et al., 2017).

8.4.2 False Positive Rates, False Negative Rates and DET Plots

Often the value 1 for the likelihood ratio (or 0 for the logarithm of the likelihood ratio) is chosen as a threshold for the proposition which is supported by the likelihood ratio. Thus

(a) for comparisons, false negative and false positive rates can be estimated. These are the number of same-source comparisons with LR < 1 divided by the total number of

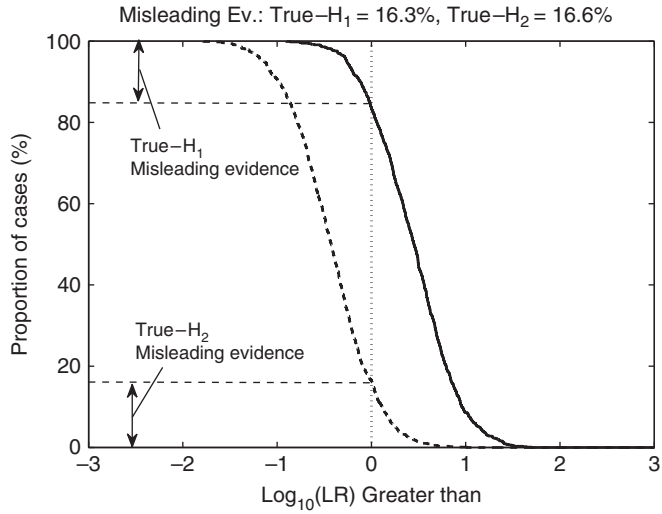

Figure 8.3 Tippett plots of a validation set of LR values artificially generated. The curve for true H_p values is the solid line on the right and the curve for the true H_d values is the dashed line on the left. The rates of misleading evidence are indicated in the title of the plot. Source: From Zadora et al. (2014), John Wiley & Sons.

same-source comparisons (*false negatives*) and the number of different-source comparisons with LR > 1 divided by the total number of different-source comparisons (*false positives*);

(b) for discrimination, the number of allocations of members of the validation set that are to the wrong group, divided by the total number of allocations provide error rates. For discrimination between two groups, A and B, say, two rates can be determined, first, the number of members of A allocated to B divided by the total number of members of A and, second, the number of members of B allocated to A divided by the total number of members of B. If there are more than two groups, various possible combinations of rates may be calculated.

However, it may be that a different threshold, τ, say, for the logarithm of the likelihood ratio is chosen. A criterion for the choice of threshold could be the one that provides the optimal values of false positive and false negative rates. Note that as the rate of false positives changes in one direction (positive or negative) with a change in threshold the rate of false negatives changes in the other direction (negative or positive). A plot of false positives *versus* false negatives as τ changes is known as a *detection error trade-off* plot or DET plot. Examples of DET plots are shown in Figure 8.7.

8.4.3 Empirical Cross-Entropy

Another measure of performance for comparisons is the so-called *log-likelihood ratio cost, Cllr*. Let n_p be the number of same-source comparisons and n_d be the number of different-source comparisons in a test of performance. The values of the likelihood ratios calculated for the same-source comparisons are denoted $\mathrm{LR}_{pi}, i = 1, \ldots, n_p$ and for the different-source comparisons are denoted $\mathrm{LR}_{di}, i = 1, \ldots, n_d$. The Cllr gives higher penalties for larger values of misleading evidence (van Leeuwen and Brummer, 2007).

It is defined as

$$\mathrm{Cllr} = \frac{1}{2}\left(\frac{1}{n_p}\sum_{i=1}^{n_p}\log_2\left(1+\frac{1}{\mathrm{LR}_{pi}}\right) + \frac{1}{n_d}\sum_{i=1}^{n_d}\log_2(1+\mathrm{LR}_{di})\right),$$

where the prior probabilities of the two propositions of same source (H_p) or different source (H_d) are taken to be equal: $\Pr(H_p) = \Pr(H_d) = 0.5$. The expression can also be written as

$$\mathrm{Cllr} = -\frac{1}{2}\left(\frac{1}{n_p}\sum_{i=1}^{n_p}\log_2(\Pr(H_p \mid E)) + \frac{1}{n_d}\sum_{i=1}^{n_d}\log_2(\Pr(H_d \mid E))\right).$$

Note that values of $\mathrm{LR}_{pi} < 1$ and $\mathrm{LR}_{di} > 1$ lead to increases in Cllr.

A variation on this measure of performance to account for different probabilities for these propositions is given in Ramos-Castro et al. (2013) and Lucena-Molina et al. (2015b). Further details of measures of performance including descriptions of *strictly proper scoring rules* and *calibration* may be found in Ramos-Castro and Gonzalez-Rodriguez (2013) and Ramos-Castro et al. (2017).

The score used by Lindley (1991) was the quadratic scoring rule (1.4). Another scoring rule is the logarithmic rule (Good, 1952).

In the context of forensic science, consider the prosecution and defence propositions H_p and H_d, respectively, and in this context, assume $\Pr(H_p) = 1 - \Pr(H_d)$. For evidence evaluation, the logarithmic rule with base 2 is used. The base of the logarithm is irrelevant for the theory and application. The base 2 is normally used in this context for reasons associated with information theory where the common unit of information is the bit, based on logarithms to base 2. Given a particular model, let p be the posterior probability obtained for H_p given evidence E and background information I. Then $(1 - p)$ is the posterior probability for H_d given evidence E and background information I. Determination of a posterior probability requires knowledge of a prior probability. The logarithmic scoring rule states that

- If H_p is true, the score is $-\log_2 p = -\log_2 \Pr(H_p \mid E, I)$;
- If H_d is true, the score is $-\log_2(1 - p) = -\log_2 \Pr(H_d \mid E, I)$.

If p is high and H_p is true then the score is low. If p is high and H_d is true then the score is high. For example, consider $p = 0.9$ and H_p true; the score is $-\log_2(0.9) = +0.15$. If H_d true; the score is $-\log_2(1 - 0.9) = +3.32$.[4] See Meuwly et al. (2017) and Ramos-Castro and Gonzalez-Rodriguez (2013) for further details.

There are two propositions. In a comparison problem the prosecution proposition H_p is that the control and recovered evidence come from the same source. The defence proposition H_d is that the control and recovered evidence come from different sources. Let $O(H_p \mid I) = \Pr(H_p \mid I)/\Pr(H_d \mid I)$. If H_p and H_d are mutually exclusive and exhaustive, then $O(H_p \mid I)$ is the odds in favour of H_p. Then, for evidence E, Bayes' theorem shows that

$$\begin{aligned} O(H_p \mid E, I) &= \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)} \times O(H_p \mid I) \\ &= V \times O(H_p \mid I), \end{aligned}$$

where $O(H_p \mid E, I) = \Pr(H_p \mid E, I)/\Pr(H_d \mid E, I)$ and $V = \Pr(E \mid H_p, I)/\Pr(E \mid H_d, I)$, the likelihood ratio. When H_p and H_d are complementary propositions, so that $\Pr(H_p) = 1 - \Pr(H_d)$, it can be shown that

$$\Pr(H_p \mid E, I) = \frac{V \times O(H_p \mid I)}{1 + V \times O(H_p \mid I)}.$$

Consider a set $\{E_{ci}, E_{ri}, i = 1, \ldots, n_p\}$ of evidential pairs (control and recovered) for same-source

[4]Note for calculation purposes, $\log_2 x = \log_{10}(x)/\log_{10}(2)$.

comparisons and a set $\{E_{cj}, E_{rj}, i = 1, \ldots, n_d\}$ of evidence pairs (control and recovered) for different-source comparisons.

The overall measure of performance for a method of evaluation of evidence may be defined as the mean value of the logarithmic scoring rule over many different comparisons of control and recovered evidence from a validation set for which the outcome of the comparison, same-source or different-source, is known. This mean is known as the *logarithmic loss*, L:

$$L = \frac{1}{n_p} \sum_{i=1}^{n_p} \log_2 \Pr(H_p \mid E_{ci}, E_{ri}, I) - \frac{1}{n_d} \sum_{j=1}^{n_d} \log_2 \Pr(H_d \mid E_{cj}, E_{rj}, I).$$

The measure of performance for evidence evaluation is then a weighted average value of the logarithmic scoring rule, and is known as the *empirical cross-entropy* (ECE):

$$\mathrm{ECE} = -\frac{\Pr(H_p \mid I)}{n_p} \sum_{i=1}^{n_p} \log_2 \Pr(H_p \mid E_{ci}, E_{ri}, I) - \frac{\Pr(H_d \mid I)}{n_d} \sum_{j=1}^{n_d} \log_2 \Pr(H_d \mid E_{cj}, E_{rj}, I)$$

$$= \frac{\Pr(H_p \mid I)}{n_p} \sum_{i=1}^{n_p} \log_2 \left(1 + \frac{1}{V_i \times O(H_p \mid I)}\right) + \frac{\Pr(H_d \mid I)}{n_d} \sum_{j=1}^{n_d} \log_2 (1 + V_j \times O(H_p \mid I)), \tag{8.7}$$

where

$$V_i = \frac{\Pr(E_{ci}, E_{ri} \mid H_p, I)}{\Pr(E_{ci}, E_{ri} \mid H_d, I)},$$
$$V_j = \frac{\Pr(E_{cj}, E_{rj} \mid H_p, I)}{\Pr(E_{cj}, E_{rj} \mid H_d, I)}. \tag{8.8}$$

This measure indicates better performance when the likelihood ratio leads to the correct decision. The numerical value will be lower as the performance increases. Since the prior odds are not known, it is not possible to evaluate a particular value of the ECE. The ECE can be represented as an ECE-plot, showing its value for a certain range of priors (Ramos-Castro et al., 2013).

Consider an increase in information to mean a reduction in uncertainty about $\Pr(H_p)$ and $\Pr(H_d)$. If values of the likelihood ratio given by the model under consideration are increasingly misleading to the fact-finder then ECE values increase; more information is needed to know which proposition,

H_p or H_d, is true. If values of the likelihood ratio given by the model under consideration increasingly support the correct proposition, then ECE values decrease.

Three curves illustrate the effectiveness of the model under consideration. The first is the curve for which the LR equals 1 for all prior odds. The posterior odds equal the prior odds. The second curve plots the ECE against the prior odds. It is to be hoped that the second curve is always below the first curve. The third curve represents the result for a model that is perfectly calibrated in that it gives the correct answer each time. This curve can only be obtained with the use of a training set or validation set and is obtained using the so-called *pool adjacent violators algorithm*. It is to be hoped the second curve is close to this curve.

8.4.3.1 Pool Adjacent Violators Algorithm

Consider a variable H that has two realisations, propositions denoted H_p and H_d so $H = \{H_p, H_d\}$. For example, in a comparison problem with control and recovered evidence E_c and E_r, respectively, the propositions may refer to the same source (H_p) and different sources (H_d) for $E = \{E_c, E_r\}$.

Consider a validation set of n evidential pairs $(E_i, P_i), i = 1, \ldots n$ where $E_i = (E_{ic}, E_{ir}), i = 1, \ldots, n$. The variable P is a binary variable that takes the value 0 if H_d is true and the value 1 if H_p is true. As the dataset is a validation set, the components $P_i, i = 1, \ldots, n$ are known and are a set of zeroes and ones. Determine the likelihood ratio for members of the validation set. Choose

a prior distribution for H_p and H_d, assuming the propositions are mutually exclusive and exhaustive. Hence obtain a set of posterior probabilities y_i from the relationship:

$$y_i = \Pr(H_p \mid E, I) = \frac{\Pr(E_i \mid, H_p, I)\ \Pr(H_p \mid I)}{\Pr(E_i \mid, H_p, I)\ \Pr(H_p \mid I) + \Pr(E_i \mid, H_d, I)\ \Pr(H_d \mid I)}. \tag{8.9}$$

The y_i, ordered in increasing magnitude, are compared with the corresponding P_i using an algorithm known as the *pool adjacent violators* (PAV) algorithm.[5] Using the PAV algorithm, an example of the use of which is given in Section 8.4.3.2, the P_i are transformed to P_i^*. The P_i^* and chosen prior odds can be used to obtain a transformed likelihood ratio V_i^*. The prior odds can be varied to obtain a variable ECE using the expression

$$\mathrm{ECE} = \frac{\Pr(H_p \mid I)}{n_p} \sum_{i=1}^{n_p} \log_2 \left(1 + \frac{1}{V_i^* \times O(H_p \mid I)}\right) + \frac{\Pr(H_d \mid I)}{n_d} \sum_{j=1}^{n_d} \log_2(1 + V_j^* \times O(H_p \mid I)). \tag{8.10}$$

[5]The name 'pool adjacent violators' may arise from consideration of the $\{P_i\}$. Adjacent P_i which do not satisfy (do violate) a monotonicity requirement are pooled to provide a set $\{P_i^*\}$ which does satisfy a monotonicity requirement that as y_i increases, so the frequency of $P_i = 1$ increases.

The variation in ECE as the prior odds vary can be plotted. See Figure 8.9 for examples. This curve is described as 'calibrated accuracy' in Zadora et al. (2014) and shows the 'performance of a validation set of optimally calibrated LR values'.

As an example (Zadora et al., 2014) consider a validation set consisting of 11 pairs $\{(y_i, P_i), i = 1, \ldots, 11\}$ in Table 8.1. Use of a validation set and (8.9) gives the values $y_i = \Pr(H_p \mid E, I)$, $i = 1, \ldots, 11$, which may be thought of as scores.

Consider pair 3 in Table 8.2 for illustration. The model used for evidence evaluation determines the posterior probability $y_3 = 0.28$ from (8.9) and the evidence E is such that E_c and E_r are known, from the validation set, to come from the same source (H_p is true, $P_3 = 1$). If the knowledge about the source of the evidence was not available and comparison was based on the posterior probability then this would be a misclassification. This argument assumes that decision

Table 8.2 Eleven illustrative pairs of posterior probabilities $\Pr(H_p \mid E, I) = y$ and labels $P = 1$ if H_p is true and $P = 0$ if H_d is true

Pair i	1	2	3	4	5	6	7	8	9	10	11
y_i	0.02	0.03	0.28	0.34	0.40	0.62	0.64	0.72	0.81	0.90	0.95
P_i	0	0	1	0	1	0	0	1	0	1	1

Evidence $E = \{E_c, E_r\}$ denotes control (c) and recovered (r) data. Proposition H_p is that E_c and E_r have the same source and proposition H_d is that E_c and E_r have different sources. Background information is denoted I.

costs (utilities) are equal; an extension to include such considerations is not discussed here.

8.4.3.2 Implementation of the PAV Algorithm

Intuitively, as the underlying probability of H_p increases, y_i, the posterior probability of H_p, assigned on the basis of the validation set, should increase. The objective of the algorithm is to create an empirical distribution function satisfying this condition for the posterior probability P of H_p given $E = \{E_c, E_r\}$ and I. It is known in advance, it is a *ground truth*, to which group the evidence E_i in the validation set belongs, namely, either that of same source H_p or that of different source H_d. This knowledge is denoted $P_i (i = 1, \ldots, n)$ in Table 8.2 with $P_i = 1$ if the corresponding member of the validation set is assigned to H_p and $P_i = 0$ if the corresponding member of the validation set is assigned to H_d. The PAV algorithm transforms these assignments to probabilities $P_i^*, i = 1, \ldots, n$ with a monotonicity requirement such that $P_i^* \leq P_{i+1}^*$ for $i = 1, \ldots, (n-1)$. As y increases, so the probability $P^* = \Pr(H_p \mid E, I)$ increases.

The solution is dependent only on the order of the posterior probabilities. After sorting the input scores $y_i, i = 1, \ldots, 11$, it is only the sequence of occurrences of $P = 1$ and $P = 0$ that matters, not the values of the scores. This means it does not matter whether the y_i are likelihood-ratios, log-likelihood ratios, or probabilities. This solution works for all sortable scores, with the meaning

that larger scores favour $H_p(P = 1)$ and smaller scores favour $H_d(P = 0)$ (Brümmer, 2010).

Consider the individual P_i in order. First, $P_2 = P_1 = 0$ so $P_2^* = P_1^* = 0$. Second, $P_2 < P_3 = 1$ but $P_3 > P_4 = 0$. Replace P_3 and P_4 by their mean value $1/2$ so $P_3^* = P_4^* = 1/2$. Now $P_4^* = 1/2 < P_5 = 1$ but $P_5 > P_6 = 0$. Replace P_5 and P_6 by their mean value $1/2$ so $P_5^* = P_6^* = 1/2$. Then $P_6^* = 1/2 > P_7 = 0$. Replacement of $P_6^* = 1/2$ and $P_7 = 0$ by their mean $1/4$ gives a result with $P_5^* = 1/2$, which is greater than $1/4$ and so the sequence does not satisfy the monotonicity requirement. This problem is resolved by assigning the mean number of 1's for pairs 3–7 inclusive. Thus $P_3^* = P_4^* = P_5^* = P_6^* = P_7^* = 2/5$. For pairs 8 and 9, $P_8 = 1 > P_9 = 0$. These can be replaced by their mean value $1/2$ and still satisfy the monotonicity requirement with previous values of P^* so $P_8^* = P_9^* = 1/2$. Finally, $P_9^* = 1/2 < P_{10} = P_{11} = 1$ so $P_{10}^* = P_{11}^* = 1$. The transformed posterior probabilities P^* are then given in Table 8.3.

8.4.3.3 Transformation of ECE

Consider the following expression for ECE:

$$\mathrm{ECE} = -\frac{\Pr(H_p)}{N_1} \sum_{i:\, H_p \text{is true}} \log_2 \Pr(H_p \mid E_i, I)$$

$$-\frac{\Pr(H_d)}{N_2} \sum_{i:\, H_d \text{is true}} \log_2 \Pr(H_d \mid E_i, I).$$

Table 8.3 Eleven sets of posterior probabilities $\Pr(H_p \mid E, I) = y$ and labels $P = 1$ if H_p is true and $P = 0$ if H_d is true, transformed using the PAV algorithm to P^*, updated probabilities $\Pr(H_p \mid E, I)$ with values given as fractions for clarity

Pair i	1	2	3	4	5	6	7	8	9	10	11
y_i	0.02	0.03	0.28	0.34	0.40	0.62	0.64	0.72	0.81	0.90	0.95
P_i	0	0	1	0	1	0	0	1	0	1	1
P_i^*	0	0	2/5	2/5	2/5	2/5	2/5	1/2	1/2	1	1

Evidence $E = \{E_c, E_r\}$ denotes control (c) and recovered (r) data. Proposition H_p is that E_c and E_r have the same source. and proposition H_d is that E_c and E_r have different sources. Background information is denoted I.

For evidence E, the posterior probabilities $\Pr(H_p \mid E, I)$, and hence $\Pr(H_d \mid E, I)$ are calculated as functions of $\Pr(H_p)$ and $\Pr(H_d)$, which is $1 - \Pr(H_p)$. These probabilities are then transformed using the empirical cumulative distribution function determined from the PAV algorithm as in Table 8.3. The resultant ECE is then

$$\mathrm{ECE} = -\frac{\Pr(H_p)}{N_1} \sum_{i:\, H_p \text{is true}} \log_2 \overset{*}{\Pr}(H_p \mid E_i, I)$$
$$-\frac{\Pr(H_d)}{N_2} \sum_{i:\, H_d \text{is true}} \log_2 \overset{*}{\Pr}(H_d \mid E_i, I).$$

A plot of ECE against prior odds $\Pr(H_p \mid I)/\Pr(H_d \mid I)$ is used to illustrate the performance of a validation set of optimally calibrated LR values obtained by a transformation applied to the

original validation set of LR values using $\{\overset{*}{\Pr}\}$. This curve is not possible to obtain in practice. It represents a *ceiling of performance*[6] (Zadora et al., 2014) useful for measuring calibration.

For further details, see Ayer et al. (1955) and Ramos-Castro et al. (2013).

8.5 CASE STUDY: KINSHIP ANALYSIS

In addition to the parent–child investigation in traditional kinship analysis, other kinds of relationships of individuals also need to be tested. An example is given here of a situation involving the analysis of kinship for possible inheritance consequences. See Taroni et al. (2005) for further details. Two individuals A and B wish to know if they are full sibs or unrelated. Let H_p be the proposition that they are full sibs and H_d be the proposition that they are unrelated. Before DNA profile analyses are conducted, it is of interest to known if a value of the likelihood ratio supporting H_p or H_d can be obtained. The methodology is extended to cover half-sibs as well as full sibs and unrelated individuals. A subsequent question

[6]The ECE curve obtained using the PAV algorithm is a minimum for possible posterior probabilities so the term 'ceiling' may be thought a misnomer. The term is chosen since the curve indicates the best (highest) performance obtainable by a model for evidence evaluation.

considers the decision on whether or not to perform DNA profile analyses and a discussion of this question is in Taroni et al. (2005).

The case study here concerns the first question regarding the values of likelihood ratios. Allele proportions (at different loci) from a selected population database are chosen. Databases of large numbers of pairs of full siblings, half-siblings, and unrelated individuals are generated using methods suggested by Triggs and Buckleton (2002). Three pairs of likelihood ratios for individuals with genotypes G_A and G_B are generated.

- Full sibs *versus* unrelated:

$$\frac{\Pr(G_A, G_B \mid \text{full sibs})}{\Pr(G_A, G_B \mid \text{unrelated})}; \quad G_A, G_B \text{ full sibs.}$$

$$\frac{\Pr(G_A, G_B \mid \text{full sibs})}{\Pr(G_A, G_B \mid \text{unrelated})}; \quad G_A, G_B \text{ unrelated.}$$

- Full sibs *versus* half-sibs:

$$\frac{\Pr(G_A, G_B \mid \text{full sibs})}{\Pr(G_A, G_B \mid \text{half-sibs})}; \quad G_A, G_B \text{ full sibs.}$$

$$\frac{\Pr(G_A, G_B \mid \text{full sibs})}{\Pr(G_A, G_B \mid \text{half-sibs})}; \quad G_A, G_B \text{ half-sibs.}$$

- Half-sibs *versus* unrelated:

$$\frac{\Pr(G_A, G_B \mid \text{half-sibs})}{\Pr(G_A, G_B \mid \text{unrelated})}; \quad G_A, G_B \text{ half-sibs.}$$

$$\frac{\Pr(G_A, G_B \mid \text{half-sibs})}{\Pr(G_A, G_B \mid \text{unrelated})}; \quad G_A, G_B \text{ unrelated.}$$

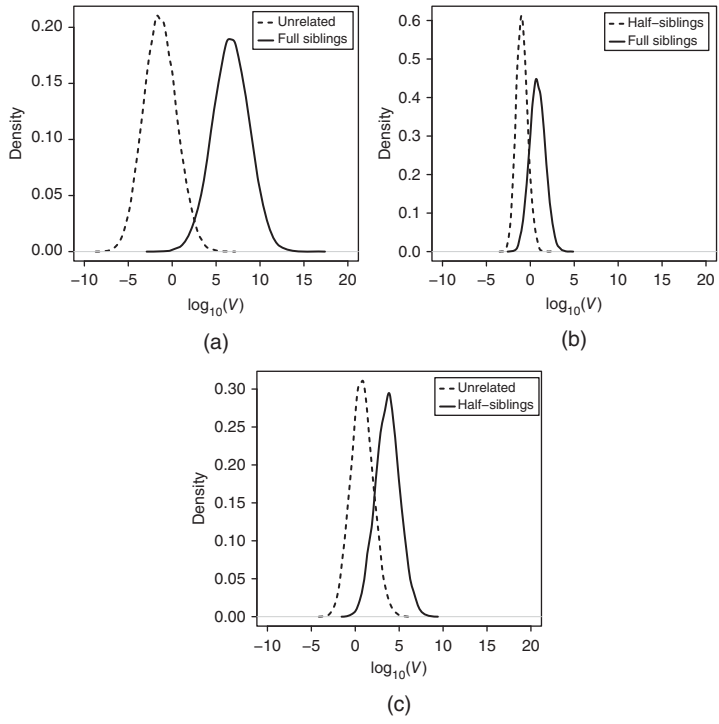

Figure 8.4 Probability density functions of likelihood ratios: (a) unrelated versus full siblings, (b) half-siblings versus full siblings, (c) unrelated versus half-siblings.

The estimates of the probability density functions of the log-likelihood ratios for the three pairs of propositions are shown in Figure 8.4. The separation for the comparison of unrelated pairs to full siblings is well marked in Figure 8.4a with little overlap of the distributions. This is in contrast to the comparison of half-siblings to full siblings and unrelated versus half-siblings where

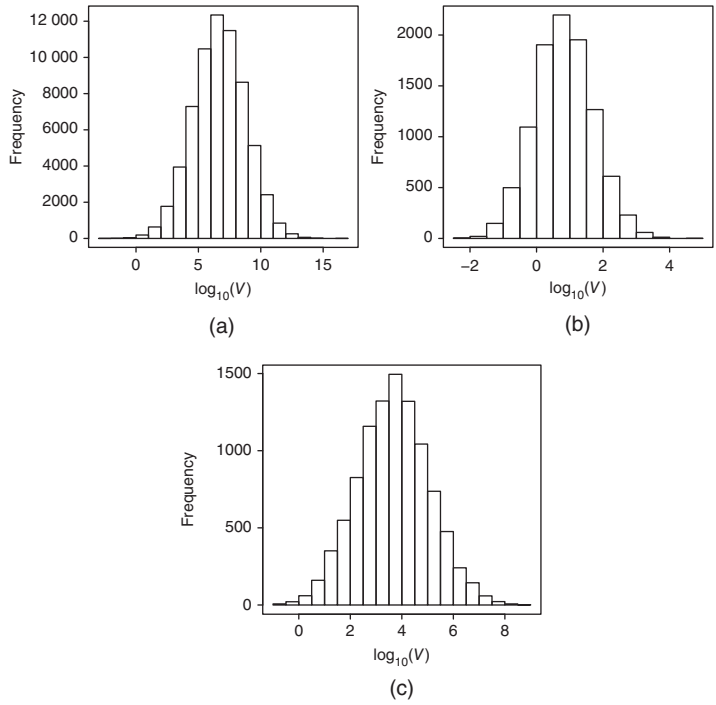

Figure 8.5 Histograms of log likelihood ratios: (a) unrelated versus full siblings assuming full siblings, (b) half-siblings versus full siblings assuming full siblings, (c) unrelated versus half-siblings assuming half-siblings.

there is considerable overlap of the distributions, see Figure 8.4b and c.

The histograms are shown in Figures 8.5 and 8.6. These are included for completeness. They are not required for large databases as used here but will be for smaller validation databases.

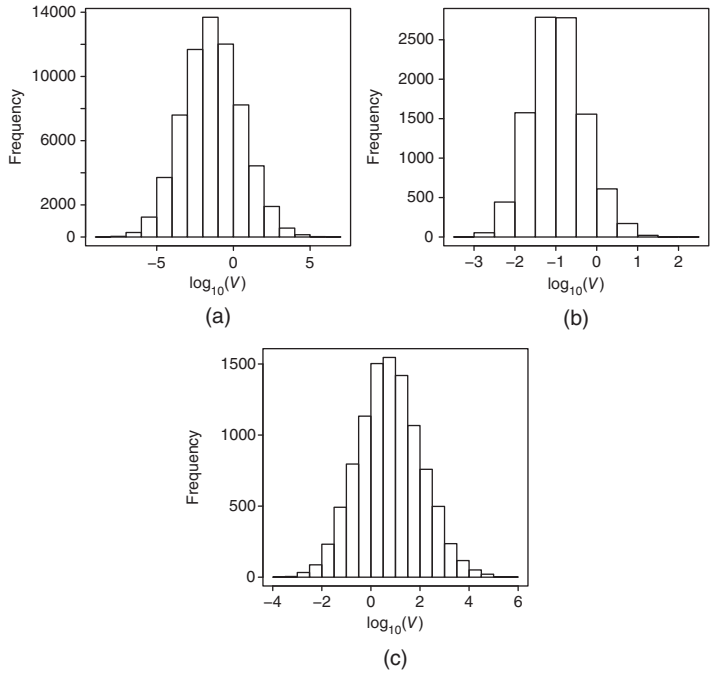

Figure 8.6 Histograms of log likelihood ratios: (a) unrelated versus full siblings assuming unrelated, (b) half-siblings versus full siblings assuming half-siblings, (c) unrelated versus half-siblings assuming unrelated.

The DET plots are shown in Figure 8.7 and the Tippett plots in Figure 8.8. A DET curve shows the relationship between the rates of false positives and false negatives. As the rate of one decreases, the rate of the other increases. The curve is of negative gradient. It is desirable that both rates are small. In such a situation, the curve will be in the bottom left-hand corner of the space. Under this criterion (a) provides the best discrimination.

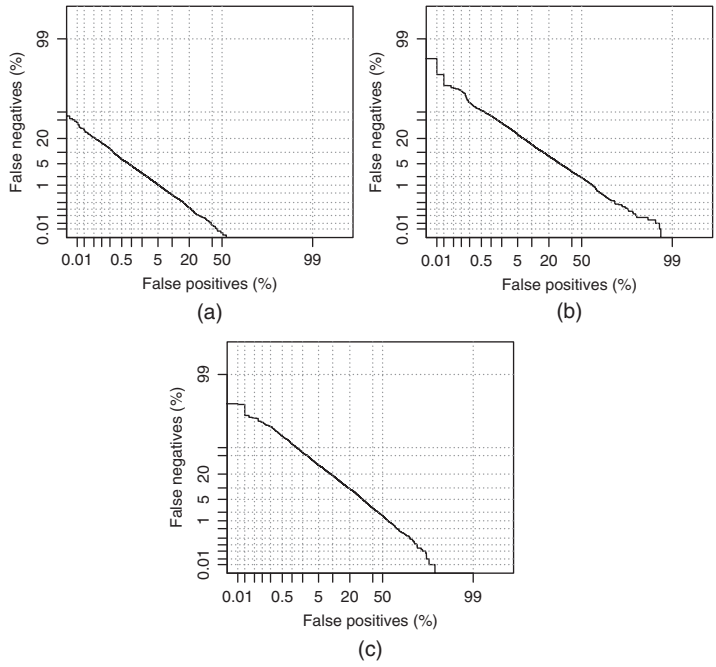

Figure 8.7 DET curves: (a) unrelated versus full siblings, (b) half-siblings versus full siblings, (c) unrelated versus half-siblings.

Relationships (b) and (c) have approximately the same effectiveness; there is little to choose between the two DET curves. The relative effectiveness of the three relationships indicated in the DET curves is mirrored in the Tippett plots. The rates of misleading evidence are small in (a), not so small in (b), and very large for unrelated misclassified as half-sibling in (c).

The ECE plots are shown in Figure 8.9. The performance of the half-sibling *versus* unrelated is very poor. The performance of the half-sibling

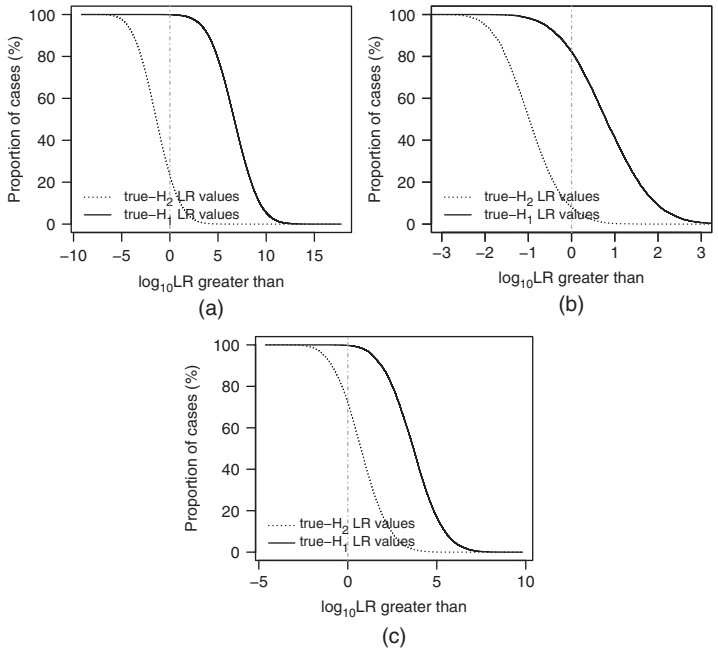

Figure 8.8 Tippett plots: (a) unrelated versus full siblings, (b) half-siblings versus full siblings, (c) unrelated versus half-siblings.

versus full sibling is excellent. This is in disagreement with the poor separation of the corresponding distributions. Ramos-Castro and Gonzalez-Rodriguez (2013) define properties of accuracy, discrimination, and calibration. For accuracy, they state that the lower the solid curve, the more accurate the method is. For discrimination, they state that the lower the dashed curve the better the discriminating power. Finally, the closer the solid and dashed curves are, the better the calibration.

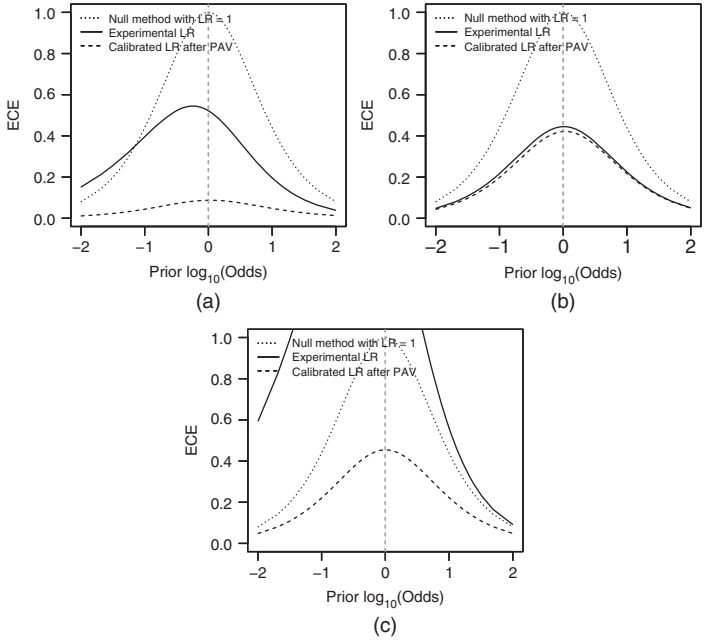

Figure 8.9 ECE plots: (a) unrelated versus full siblings; (b) half-siblings versus full siblings; (c) unrelated versus half-siblings.

8.6 CONCLUSION

The use of the criteria for the assessment of performance will be subjective. If the various criteria are deemed to be good then the method can be used. The following comments may help the judgement.

- *False positives and false negatives*: false positives are more undesirable than false negatives in a comparison problem. Both rates have to be low for a method to be usable but how low is a subjective decision.

- *Tippett plots*: again it is a subjective decision as to how close the plots are to perfection. The plot for the prosecution proposition, corresponding to the numerator in the likelihood ratio, should be close to 1 for the likelihood ratio less than 1 (log likelihood ratio less than 0). The plot for the defence proposition, corresponding to the denominator in the likelihood ratio, should be close to 0 for the likelihood ratio greater than 1 (log likelihood ratio greater than 0).
- *ECE plots*: the data curve has to be close to the one obtained from the PAV algorithm and below the one assuming a likelihood ratio of 1 throughout. If the data curve crosses the 1 assuming a likelihood ratio of 1 in a big way (with a subjective decision as to what is meant by 'big'), then the method should not be used, though this stricture may only apply in cases where the log prior odds takes values where the data curve crosses the curve for a likelihood ratio of 1.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

A

Probability Distributions

A.1 INTRODUCTION

Various probability distributions have been mentioned in the course of the book. They are summarised here in the Appendix for ease of reference with a few examples to aid understanding. In certain general circumstances the way in which probability is distributed over the possible numbers of counts or values for the measurements can be represented mathematically, by functions known as probability distributions. Distributions for counts and for measurements will be described in Sections A.2 and A.3, respectively. Further details of many of the distributions mentioned here and others may be found in Forbes et al. (2010) or in Wikipedia. Before probability distributions can be discussed here, however, certain other concepts have to be introduced.

A characteristic of interest from a population is known as a *parameter*. The corresponding characteristic from a sample from the population is known as an *estimate*. For example, the proportion γ_{AA} of Caucasians in Chicago with allele AA at locus *LDLR* is a parameter. The proportion of Caucasians with allele AA at locus *LDLR* in the sample of 200 people studied by Johnson and Peterson (1999) is an estimate of γ_{AA}. Conventionally a ˆ symbol (read as 'hat') is used to denote an estimate. Thus $\hat{\gamma}_{AA}$ (read as 'gamma-hat AA') denotes an estimate of γ_{AA}. From Table 1.1, $\hat{\gamma}_{AA} = 0.188$.

Another notational convention, as well as the ˆ notation, is to use Roman letters for functions evaluated from counts or measurements from samples and Greek letters for the corresponding parameters from populations. Thus given a sample of size n from a population with measurements $\{x_1, \ldots, x_n\}$, the sample mean is $(x_1 + \cdots + x_n)/n$, which may be summarised as $\sum_{i=1}^{n} x_i/n$ and denoted $\bar{x}$ (read as 'x-bar'). The corresponding population mean is denoted with a Greek letter, e.g. μ. A sample standard deviation may be denoted s and for the sample $\{x_1, \ldots, x_n\}$ an estimate of the standard deviation is the square root of $\sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)$. The corresponding population standard deviation is often denoted σ. The square of the standard deviation is known as the *variance*; a sample variance may be denoted s^2, and the corresponding population variance σ^2.

The concept of a *random variable* (or *random quantity* or *uncertain quantity*) (Lindley, 1991) needs some explanation also. A *random variable*, in a rather circular definition, is something that varies at random. For example, the number of sixes in rolls of four dice varies randomly amongst the five numbers {0, 1, 2, 3, 4} as the dice are rolled for several sets of rolls; the number of sixes in each set of rolls is a discrete random variable. Similarly, the refractive index of a fragment of glass varies over the set of all fragments of glass and is a continuous random variable. The variation to be considered in the refractive index of glass is however of a more complicated structure than the number of sixes in rolls of four dice. There is variation in refractive index within a window and between windows. This requires parameters to measure two standard deviations, one for each type of variation, and this problem is discussed in greater detail in Chapter 7.

Notation is useful in the discussion of random variables. Rather than write out in long-hand phrases such as 'the number of sixes in rolls of four dice' or 'the refractive index of a fragment of glass', the phrases may be abbreviated to a single upper-case Roman letter. For example, let X be short for 'the number of sixes in rolls of four dice'. It then makes sense to write mathematically $\Pr(X = 3)$, which may be read as 'the probability that the number of sixes in rolls of four dice equals 3'. More generally still, the 3 may be replaced by a

lower-case Roman letter to give $\Pr(X = x)$, say, where x may then be substituted by one of the permissible values $\{0, 1, 2, 3, 4\}$ as required.

Similarly, X may be substituted for 'the refractive index of a fragment of glass' and the phrase 'the probability that the refractive index of a fragment of glass is less than 1.5185' may be written as $\Pr(X < 1.5185)$, or more generally as $\Pr(X < x)$ for a general value x of the refractive index. For reasons that are explained later (Section A.3.2), it is not possible to evaluate $\Pr(X = x)$ for a random variable representing a continuous measurement. In general, upper case Roman letters (such as X or Y) represent a random variable and lower case Roman letters (such as x or y) represent a particular value (sometimes known as a *realisation*) of a random variable.

The mean of a random variable is the corresponding population mean. In the examples earlier this would be the mean number of sixes in the conceptual population of all possible sets of rolls of four dice (and here note that the population need not necessarily exist except as a concept) or the mean refractive index of the population of all fragments of glass (again, effectively, a conceptual population). The mean of a random variable is given a special name, the *expectation*, and for a random variable, X, say, it is denoted $\mathrm{E}(X)$. Similarly, the variance of a random variable is the corresponding population variance. For a random variable X, it is denoted $\mathrm{Var}(X)$.

Observations $x_1, \ldots, x_n$ from a random sample[1] from a population may be treated as realisations of random variables $X_1, \ldots, X_n$. A *statistic* is a function of the data. The function of the corresponding random variable has a distribution. For example, $\bar{x} = \sum_{i=1}^{n} x_i/n$ is a sample mean. The corresponding random variable may be denoted $\bar{X} = \sum_{i=1}^{n} X_i/n$. Thus, the sample mean and the sample variance are statistics. A particular value of a statistic that is determined to estimate the value of a parameter is known as an *estimate*. The corresponding random variable is known as an *estimator*. An estimator, X, say, of a parameter, θ, say, which is such that $\mathrm{E}(X) = \theta$ is said to be unbiased. If $\mathrm{E}(X) \neq \theta$, the estimator is said to be biased.

It is hoped that an estimate will be a good estimate in some sense. Different samples from the same population may produce different estimates. The proportion of people with an AA allele at locus *LDLR* in a second sample of 200 Caucasians from Chicago may have produced a different number of people with AA alleles from the first example and hence a different value for $\hat{\gamma}_{\mathrm{AA}}$. Different results from different samples do not mean that some are wrong and others are right. They merely indicate the natural variability in the distribution of allelic frequencies amongst people.

Properties of a procedure for estimation of a parameter are given with reference to the

[1] A random sample from a population is one in which every member of the population has an equal probability of selection.

estimator. The estimator is a random variable and has an expectation and variance. An estimator may be considered good if it is *accurate* and *precise* (Section 8.2). Accuracy is a measure of closeness of an estimator to the true value of the parameter of interest. For example, this closeness may be the absolute difference between the expectation and the true value of the parameter. Precision is related to the variance of the estimator. In mathematical statistical terminology, precision is synonymous with the reciprocal of the variance. The greater the variance of an estimator, the less its precision.

Once a sample has been selected, the particular value of the estimator applicable to that sample is known as an estimate. If different samples lead to estimates with different sample means and sample variances of the same characteristic, then there is a suggestion that the variability is great and the estimates are not very precise. For example, if the variability in the estimation procedure is large then a second estimate, from a different sample of people than those in Table 1.1, of γ_{AA} may produce an estimate very different from 0.188 ($\hat{\gamma}_{AA}$).

In the example earlier it is obviously desirable that $\hat{\gamma}_{AA}$ be close to γ_{AA}. Precision is a measure of the variability of the estimators, whether or not they are close to the true value (Dodge, 2006). It is possible to have an accurate estimator that is not very precise and a precise estimator that is not very accurate.

The importance of allowing for variability is illustrated by the following hypothetical example

from a medical context. The reaction times of two groups of people, group *A* and group *B*, say, are measured. Both groups have the same median[2] reaction time, 0.20 seconds, but group *A*'s times vary from 0.10 to 0.30 seconds, whereas group *B*'s range from 0.15 to 0.25 seconds. Samples of equal numbers of people from each group are then given a drug designed to reduce reaction times. In both cases, the reaction times of the samples of people given the drug range from 0.11 to 0.14 seconds. For group *A*, this is within the range of previous knowledge and there is a little, but not very strong, evidence to suggest that the drug is effective in reducing reaction times. For group *B*, however, the result is outwith the range of previous knowledge and there is very strong evidence to suggest that the drug is effective. Note that both group *A* and group *B* had the same initial median reaction time. The drug produced the same range of reaction times in samples from both groups. The distinction in the interpretation of the results between the two groups arises because of the difference in the range, or variability, of the results for the whole of each of the two groups.

Later, in Section A.3.2, it will be seen that when measurements are standardised, variation is accounted for by inclusion of the standard deviation.

[2]A median is the value such that 50% of the distribution is less than it and 50% greater than it; see Section A.3.1.

The applications of these concepts are now discussed in the context of probability distributions for counts and for measurements.

A.2 PROBABILITY DISTRIBUTIONS FOR COUNTS

A.2.1 Probabilities

Suppose four fair six-sided dice are to be rolled, once each. Let the event of interest be the number of occurrences of a six being uppermost; denote this by X. Then X can take one of five different integer values, 0, 1, 2, 3, or 4. Over a sequence of groups of rolls of the four dice, X will vary *randomly* over this set of 5 integers. Outcomes of successive groups of rolls are independent. For any one group of rolls of the four dice, X takes a particular value, one of the integers $\{0, 1, 2, 3, 4\}$. Denote this particular value by x.

There is a formula that enables this probability to be evaluated easily. Notice that in the roll of any one die, the probability of throwing a six is $1/6$, since the dice are fair and each side has a probability of $1/6$ of landing uppermost. The probability of not throwing a six is $5/6$, as this is a complementary event to the throwing of a six. Then, for the rolls of four dice

$$\Pr(X = x) = \binom{4}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{4-x},$$
$$x = 0, 1, \ldots, 4;$$

an example of the binomial distribution (Section A.2.3). The term $(1/6)^x$ corresponds to the x sixes, each occurring independently with probability 1/6. The term $(5/6)^{4-x}$ corresponds to the $(4\text{-}x)$ non-sixes, each occurring independently with probability 5/6. The term $\binom{4}{x}$ is the *binomial coefficient*

$$\binom{4}{x} = \frac{4!}{x!(4-x)!},$$

where $x! = x(x-1)(x-2)\cdots 1$, known as x-factorial and, conventionally, $0! = 1$. The binomial coefficient here is the number of ways in which x sixes and $(4-x)$ non-sixes may be selected from the rolls of the four dice, without attention being paid to the order in which the sixes occur.

Suppose $x = 1$, there is one six and three non-sixes, then

$$\Pr(X = 1) = \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3.$$

Now

$$\binom{4}{1} = \frac{4!}{1!3!} = \frac{4\times 3\times 2\times 1}{1\times 3\times 2\times 1} = 4,$$

$$\left(\frac{1}{6}\right)^1 = \frac{1}{6},$$

$$\left(\frac{5}{6}\right)^3 = \frac{125}{216},$$

$$\Pr(X = 1) = 4\times\frac{1}{6}\times\frac{125}{216} = 0.3858.$$

The probabilities for the five possible outcomes relating to the number of sixes in the rolls of the four dice are given in Table A.1.

Table A.1 Probabilities for the number of sixes, X, in rolls of four fair six-sided dice.

Number of sixes (x)	0	1	2	3	4	Total
$\Pr(X = x)$	0.4823	0.3858	0.1157	0.0154	0.0008	1.0000

Notice that the sum of the probabilities is 1 since the five possible outcomes 0, 1, 2, 3, and 4 are mutually exclusive and exhaustive (Sections 1.7.8 and 1.7.10).

A.2.2 Summary Measures

It is possible to determine a value for the mean of the number of sixes in rolls of four dice; this is the expectation (see Section A.1) of the number of sixes in rolls of the dice. Consider 10 000 groups of rolls of four dice. The probabilities in Table A.1 may be considered as the expected proportion of times in which each of 0, 1, 2, 3, and 4 sixes would occur. Thus it would be expected that on 4823 times there would be 0 sixes, 3858 times 1 six, 1157 times 2 sixes, 154 times 3 sixes, and on 8 times there would be 4 sixes. The total number of sixes expected is thus

$$(0 \times 4823) + (1 \times 3858) + (2 \times 1157) + (3 \times 154) + (4 \times 8) = 6666.$$

In any one group of rolls, the expected number $\mathrm{E}(X)$ of sixes is then $6666/10\,000 = 0.6666$.

Notice that this is not an achievable number (0, 1, 2, 3, or 4) but is justified by the calculations. (In a similar way, an average family size of 2.4 children is not an achievable family size.) There is a formula for the calculation of its expectation

$$\begin{aligned}\mathrm{E}(X) &= 0 \times \Pr(X = 0) + \cdots + 4 \times \Pr(X = 4) \\ &= \sum_{x=0}^{4} x \Pr(X = x). \qquad \text{(A.1)}\end{aligned}$$

This can be further shortened by denoting $\Pr(X = x)$ by θ_x so that

$$\mathrm{E}(X) = \sum_{x=0}^{4} x\,\theta_x, \text{ with } \theta_0 + \cdots + \theta_4 = 1.$$

Note that

$$\mathrm{E}(X/n) = \sum_{x=0}^{4} \frac{x}{n}\,\theta_x = \frac{1}{n}\sum_{x=0}^{4} x\,\theta_x = \mathrm{E}(X)/n.$$

In general, for $(n + 1)$ outcomes $\{0, 1, \ldots, n\}$ with associated probabilities $\theta_0, \theta_1, \ldots, \theta_n$,

$$\mathrm{E}(X) = \sum_{x=0}^{n} x\,\theta_x, \text{ with } \theta_0 + \cdots + \theta_n = 1.$$

Note the use of the Greek capital letter Σ to denote summation (*S* for *summation*). The expression below the symbol (when it is displayed) or as a subscript (when in the body of the text) denotes the term over which the summation is being made and the starting point of the sum. The finishing point of the sum is above the symbol (when it is displayed) or as a superscript (when in the body

of the text). This symbol should be compared with the Greek capital letter $\prod$ to denote product (P for *product*) where the same convention for indexing the product is used. The first example of the use of $\prod$ is in Section 3.5.4.

The expectation is a well-known statistic. Not so well-known is the *variance* that measures the variation in a set of observations. The number of sixes which occurs in any group of rolls of the four dice varies from group to group over the integers 0, 1, 2, 3, and 4.

Consider the square of the difference, $d(x)^2 = \{x - \mathrm{E}(X)\}^2$ between an outcome x and its expectation. This squared difference has a corresponding random variable $d(X)^2 = \{X - \mathrm{E}(X)\}^2$ and as such has an expectation. The expectation of $d(X)^2$, $\mathrm{E}\{d(X)^2\}$, for a set of $(n+1)$ outcomes $\{0, 1, \ldots, n\}$ with associated probabilities $\theta_0, \theta_1, \ldots, \theta_n$ is the *variance* of X, denoted $\mathrm{Var}(X)$,

$$\mathrm{Var}(X) = \sum_{x=0}^{n} \{x - \mathrm{E}(X)\}^2 \theta_x. \qquad \text{(A.2)}$$

As before, the square root of the variance is the *standard deviation.*

Note that

$$\begin{aligned}\mathrm{Var}\left(\frac{X}{n}\right) &= \sum_{x=0}^{n} \left\{\frac{x}{n} - \mathrm{E}\left(\frac{X}{n}\right)\right\}^2 \theta_x \\ &= \frac{1}{n^2} \sum_{x=0}^{n} \{x - \mathrm{E}(X)\}^2 \theta_x = \mathrm{Var}(X)/n^2.\end{aligned}$$

Another, quicker, method of evaluation of the variance is to evaluate

$$\mathrm{Var}(X) = \sum_{x=0}^{n} x^2\,\theta_x - \left(\sum_{x=0}^{n} x\,\theta_x\right)^2.$$

The variance may be determined for the example of the number of sixes in rolls of four dice as follows, where $\mathrm{E}(X) = 0.6666$.

The variance of X may then be calculated as

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x=0}^{4} \{x - \mathrm{E}(X)\}^2\,\theta_x \\ &= \sum_{x=0}^{4} d(x)^2\,\theta_x = 0.5557.\end{aligned}$$

The quicker way is to evaluate

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x=0}^{4} x^2\,\theta_x - \left(\sum_{x=0}^{4} x\,\theta_x\right)^2 \\ &= 1.0000 - 0.6666^2 = 0.5556.\end{aligned}$$

The intermediate calculations are given in Table A.2.

This example of one roll of each of four fair six-sided dice may be generalised. Consider each roll of a die as a *trial*, in a statistical context. Such a trial is different from a trial in a legal context. A statistical trial is an experiment to investigate the relationships amongst characteristics

Table A.2 Intermediate calculations for the variance of the number of sixes, x, in one roll of each of four fair six-sided dice.

x	0	1	2	3	4
d	−0.6666	0.3334	1.3334	2.3334	3.3334
d^2	0.4444	0.1112	1.7780	5.4448	11.1116
θ_x	0.4823	0.3858	0.1157	0.0154	0.0008
x^2	0	1	4	9	16

through observations of the characteristics on individuals in a random sample from a population. An example is that of the relationship amongst elemental concentrations of fragments of window glass. The population could be that of window glass, the sample would be a subset of fragments of glass from the population.

Consider the example of one roll of a fair six-sided die. The roll of each die may be considered as a trial. For each trial there will be one of only two outcomes, a six or a non-six (1, 2, 3, 4, 5). Conventionally, in general terms, these may be known as *success* (a six) and *failure* (a non-six). The trials are independent of each other. The probability of each of the outcomes is constant from trial to trial (the probability of a six is 1/6 for each die). Such a set of trials is known as a set of Bernoulli trials (after the Swiss mathematician, James Bernoulli, 1654–1705). The conditions to be met are

- fixed number of trials;

- independent trials;
- two and only two outcomes, conventionally denoted *success* and *failure* or *positive* and *negative*;
- constant probability of success from trial to trial.

A.2.3 Binomial Distribution

For the binomial distribution, and the multinomial distribution to follow, the probability of a particular outcome in any one trial is assumed constant. Thus the probability of a six in a throw of a fair die is assumed equal to 1/6 regardless of the number of throws of the die. A simplistic approach to the evaluation of a DNA profile assumes the probability of a particular allelic type is assumed constant, regardless of the number of other people who have been observed with or without that type. See Section 6.1.1 for a relaxation of this assumption. The population from which these observations have been taken (all throws of a fair die, all people in the population) is sufficiently large that the observation of a particular outcome does not alter the probability of that outcome in future trials. In a sense, it may be considered that once that outcome from the population has been observed, it is then returned to the population and may be selected for observation again. The selection (or sampling) of observations from the population is said to be *with replacement*.

Let n denote the number of independent trials. Let X denote the number of successes. Let θ denote the probability of a success in any individual trial and let $(1-\theta)$ denote the probability of a failure in any individual trial. Denote the probability, $\Pr(X = x)$, that X, the number of successes, equals x, by θ_x; $x = 0, 1, \ldots, n$. This probability is dependent on n and θ and more correctly should be written as $\Pr(X = x \mid n, \theta)$.

The situation described earlier is a very common one. Examples include the number of heads in ten tosses of a fair coin ($n = 10, \theta = 1/2$), the number of sixes in one roll of five fair dice ($n = 5, \theta = 1/6$), the number of people with genotype $\{11, 12\}$, at the *FES* locus in a sample of size 50 from a relevant population ($n = 50$, θ may be estimated from previous population data). The distribution of the probabilities (*probability distribution*) over the set of possible outcomes is known as the *binomial distribution*. The function that gives the formula for the probabilities $\Pr(X = x)$ is known as a *probability function*. For the binomial distribution $\Pr(X = x \mid n, \theta)$ is given by

$$\Pr(X = x \mid n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x},$$
$$x = 0, 1, \ldots, n; \quad 0 < \theta < 1; \tag{A.3}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \tag{A.4}$$

the *binomial coefficient*. The distribution of X can be denoted in short-hand as

$$X \sim Bin(n, \theta)$$

where $\sim$ is to be read as *is distributed as*, the first term n in $(,)$ denotes the number of trials and the second term θ denotes the probability of success. For example, if X is the number of sixes in one throw of each of 10 fair dice, then this can be denoted as

$$X \sim Bin(10, 1/6).$$

It can be shown that

$$\mathrm{E}(X) = n\theta, \ \mathrm{Var}(X) = n\theta(1-\theta).$$

(Verification of these formulae can be made by reference to the numerical results in Section A.2.2.) Note that $\mathrm{E}(X/n) = \mathrm{E}(X)/n = n\theta/n = \theta$. Thus, X/n, the sample proportion of successes, is an unbiased estimator of θ, the probability of success in an individual trial. Note also that $\mathrm{Var}(X/n) = \mathrm{Var}(X)/n^2 = \theta(1-\theta)/n$.

A.2.4 Multinomial Distribution

The multinomial distribution is a generalisation of the binomial distribution. The binomial distribution models a situation in which there is a sequence of independent trials in each of which there are only two possible mutually exclusive outcomes. The multinomial distribution models a situation in which there is a sequence of independent trials in each of which there are k possible mutually exclusive outcomes ($k \geq 2$). Denote the probabilities for the k outcomes $\theta_1, \ldots, \theta_k$ with $\sum_{i=1}^{k} \theta_i = 1$. Consider n independent trials in which the observed number of occurrences of each of the k outcomes is $x_1, \ldots, x_k$ with $\sum_{i=1}^{k} x_i = n$.

The corresponding random variables are denoted $X_1, \ldots, X_k$ where X_i is shorthand for the phrase 'the number of occurrences of outcome i'. The probability of observing $\{X_1 = x_1, \ldots, X_k = x_k\}$ is then

$$\begin{aligned}\Pr(X_1 = x_1, \ldots, X_k = x_k \mid n, \theta_1, \ldots, \theta_k) \\ = \frac{n!}{x_1!x_2!\ldots x_k!}\theta_1^{x_1}\ldots\theta_k^{x_k},\end{aligned}$$

where $\sum_{i=1}^{k} x_i = n$, $\sum_{i=1}^{k} \theta_i = 1$. This distribution may be used to model allele frequencies at loci where there are more than two possible alleles and to model drug frequencies in consignments of tablets in which there are more than two possible drug types. When $k = 3$ and there are three mutually exclusive outcomes, the distribution is also known as a *trinomial* distribution. When $k = 2$, the multinomial distribution is the binomial distribution.

A.2.5 Hypergeometric Distribution

There are instances when the population is not large and the observation of the outcome of a particular trial does change the probability of that outcome in future. For example, consider sampling from a consignment of N white tablets to determine the proportion that are illicit, an example for which further details were given in Section 4.3.2. The tablets are assumed indistinguishable by size, colour, weight, and texture

but each is assumed to be either licit or illicit. A sample of size m is taken from the consignment. The true, but unknown, number of illicit tablets is R and the true, but unknown, number of licit tablets is $N - R$. The first tablet examined is either illicit (with probability R/N, the proportion of illicit tablets in the consignment) or licit (with probability $(N - R)/N$). After examination it is put to one side; it is not placed back in the consignment. A second tablet is then examined.

Assume the first tablet was illicit. The second tablet is either illicit or licit. The probability that it is illicit is $(R - 1)/(N - 1)$, the proportion of illicit tablets remaining in the consignment. The probability that it is licit is $(N - R)/(N - 1)$, the proportion of licit tablets remaining in the consignment.

Assume the first tablet was licit. The second tablet is either illicit or licit. The probability that it is illicit is $R/(N - 1)$, the proportion of illicit tablets remaining in the consignment. The probability that it is licit is $(N - R - 1)/(N - 1)$, the proportion of licit tablets remaining in the consignment.

After examination, the second tablet is put to one side. A third tablet is examined. There are three possibilities for the probability it is illicit. These depend on the outcomes of the first two examinations, in which there may zero, one, or two illicit tablets.

Sampling in this context where N, the consignment size, is small is said to be *without replacement*. The distribution that models the probability of the

number X of illicit tablets in a sample of size n from a consignment of size N in which R are illicit and $(N - R)$ are licit is the hypergeometric distribution. The hypergeometric distribution arises in a discussion of exchangeability (Section 1.7.7).

The probability function is given by

$$\Pr(X = x \mid R, N, m) = \frac{\binom{R}{x}\binom{N-R}{m-x}}{\binom{N}{m}};$$

$$\max(0, m + R - N) \leq x \leq \min(m, R). \quad \text{(A.5)}$$

Further examples of the use of the hypergeometric distribution are given in Curran et al. (1998a) for glass comparisons and in ENFSI guidelines on sampling of illicit drugs for qualitative analysis (ENFSI, 2016).

It is possible to extend the hypergeometric distribution to the situation where there are more than two categories. This is analogous to the extension of the binomial to the multinomial. No further discussion of this extension is given here.

A.2.6 Poisson Distribution

This probability distribution is named after the French mathematician, S.D. Poisson (1781–1840). The distribution is generally used to describe the number of events that occur randomly in a specified period of time or interval of space. It is parameterised by a single parameter,

λ, say, the mean or expectation of the distribution (in unit time or space). Then, the probability the number of events X equals x in unit time or space is given by

$$\Pr(X = x \mid \lambda) = \frac{\lambda^x}{x!}\exp(-\lambda); \quad x = 0, 1, \ldots; \ \lambda > 0,$$

where $\exp\{\cdots\}$ denotes e, the base of Napierian logarithms ($e = 2.718\ 281\ 828\ldots$) and $\exp(-\lambda)$ denotes $e^{-\lambda}$. A characteristic of the Poisson distribution is that the variance equals the mean; $\mathrm{Var}(X) = \mathrm{E}(X) = \lambda$. The distribution of X can be denoted in shorthand as $X \sim Po(\lambda)$.

The parameter λ is then multiplied by the period of time or interval of space under consideration to give the mean number of events within that period. As an example in time, consider the emission of radioactive particles from a radioactive source, as measured by a Geiger counter. Take the unit of time to be one second. Denote the mean number of particles emitted in one second by λ, a number, not necessarily an integer, greater than 0. The mean number of particles emitted in t seconds is then λt, where $0 < t < \infty$. As an example in space, consider the number of characteristics of a particular kind in a piece of handwriting. Take the unit of space to be one character in the handwriting. Again, denote the mean by λ, where here this is the mean number of the particular kind of characteristic, and it would be expected that λ is very much less than 1. The mean number of characteristics of the particular kind in a length

of handwriting of s characters is then λs. Note that the parameter λ has units 'per unit time' or 'per unit interval of space'. Thus, when considering the distribution of the number of events it is important to specify the length of time or area or volume of space, which is being considered.

Consider time. Let X denote the number of events in a period of time t, which is considered as a random variable with mean λt that follows a Poisson distribution. Then, the probability that X takes a particular value x (a non-negative integer) is given by the equation

$$\Pr(X = x \mid \lambda, t) = \frac{(\lambda t)^x}{x!} \exp(-\lambda t); \quad x = 0, 1, \ldots; \lambda > 0. \tag{A.6}$$

Equation (A.6) is sometimes written, more conveniently, as

$$\Pr(X = x \mid \lambda, t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}; x = 0, 1, \ldots; \lambda > 0. \tag{A.7}$$

A.2.7 Beta-Binomial and Dirichlet-Multinomial Distributions

Consider an example of a consignment of tablets, a proportion of which are suspected to be drugs. For large consignments, the probability distribution of the proportion θ which are drugs can be modelled with a beta distribution (Sections 4.3.1 and A.3.7 later in this chapter), which treats the proportion

θ as a variable that is continuous over the interval $(0, 1)$. For small consignments, say, $N < 50$, then a more accurate distribution, which recognises the discrete nature of the possible values of the proportion, may be used (Section 4.3.2).

Of the number m sampled and inspected, z are found to be illicit. Assume there remain n units in the consignment, which are uninspected so that $m + n = N$, the total consignment size. Let Y ($\leq n$ and unknown) be the number of units in the remainder of the consignment that contain drugs. The total number of units in the consignment that contain drugs is then $(z + y)$ $(\leq N)$. The distribution for $(Y \mid m, n, z, \alpha, \beta)$ is a so-called Bayesian *predictive distribution* known as the beta-binomial distribution (Bernardo and Smith, 2000) with

$$\begin{aligned}\Pr(Y = y \mid m, n, z, \alpha, \beta) &= \frac{\begin{array}{c}\Gamma(m+\alpha+\beta)\binom{n}{y}\Gamma(y+z+\alpha)\\ \times\Gamma(m+n-z-y+\beta)\end{array}}{\begin{array}{c}\Gamma(z+\alpha)\Gamma(m-z+\beta)\\ \times\Gamma(m+n+\alpha+\beta)\end{array}};\\ y &= 0, 1, \ldots, n, \end{aligned} \tag{A.8}$$

where

$$\begin{aligned}\Gamma(x+1) &= x\Gamma(x),\\ \Gamma(x+1) &= x! \text{ for integer } x > 0,\\ \Gamma(1/2) &= \sqrt{\pi},\end{aligned}$$

is the gamma function, values of which are available from appropriate software such as R.

The derivation of this distribution requires a beta prior (Section A.3.7) and a binomial model for the data (m, z). This gives a posterior beta distribution for the proportion. This is then combined with a binomial model for the uninspected portion (n, y) of the consignment to give the beta-binomial distribution earlier. Further details are given in Section 4.3.2 and Aitken (1999).

The beta-binomial distribution may be generalised to consider more than two categories, and the corresponding distribution is known as the Dirichlet-multinomial distribution.

For the example of a consignment of tablets, there may be more than two types of drugs. For large consignments, the probability distribution of the proportions $\{\theta_i,\ i = 1, \ldots, k\}$ of the various types of drugs can be modelled with a Dirichlet distribution (Section A.3.8), which treats the proportions θ_i as variables that are continuous over the interval $(0, 1)$ with the constraint that $\Sigma_{i=1}^{k} \theta_i = 1$.

As before, consider a consignment of tablets. A sample of size m has been inspected and z_i are found to be of drug i, $i = 1, \ldots, k$ such that $\sum_{i=1}^{k} z_i = m$. Assume there are n units in that part of the consignment that has not been inspected such that $m + n = N$, the total consignment size. Let $(Y_i, i = 1, \ldots, k)$ be the numbers (unknown) of tablets in each of the k groups in the remainder of the consignment that contain drugs.

The total number of tablets in the consignment of type i is then $(z_i + y_i)$ $(\leq N)$. The distribution

for $(Y_i \mid m, n, z_1, \ldots, z_k, \alpha_1, \ldots, \alpha_k)$ is the Bayesian predictive distribution known as the Dirichlet-multinomial distribution (Bernardo and Smith, 2000) with

$$\Pr(Y_1 = y_1, \ldots, Y_k = y_k \mid m, n, z_1, \ldots, z_k, \alpha_1, \ldots, \alpha_k)$$
$$= \frac{\Gamma(m + \sum_{i=1}^{k} \alpha_i) \frac{n!}{y_1! \cdots y_k!} \prod_{i=1}^{k} \Gamma(y_i + z_i + \alpha_i)}{\prod_{i=1}^{k} \Gamma(z_i + \alpha_i)\ \Gamma(m + n + \sum_{i=1}^{k} \alpha_i)};$$
$$0 \le y_i \le n; \sum_{i=1}^{k} y_i = n. \tag{A.9}$$

The derivation of this distribution requires a Dirichlet prior (Section A.3.8) and a multinomial model (Section A.2.4) for the data $(m, z_1, \ldots, z_k)$. This gives a posterior distribution for the proportions of each of the k types. This distribution is then combined with a multinomial model for the uninspected portion $(n, y_1, \ldots, y_k)$ of the consignment to give the Dirichlet-multinomial distribution earlier. A brief further reference is given in Section 4.3.2.

A.3 MEASUREMENTS

A.3.1 Summary Statistics

Consider a population of univariate continuous measurements with parameters mean μ, variance σ^2, and standard deviation σ. These are continuous equivalents of (A.1) and (A.2) for expectation and variance of discrete random variables.

Given sample data $(x_1, x_2, \ldots, x_n)$ of measurements from this population, μ and σ may be estimated from the sample data as follows. The sample mean, denoted $\bar{x}$, is defined by

$$\bar{x} = \sum_{i=1}^{n} x_i/n. \tag{A.10}$$

The sample standard deviation, denoted s, is defined as the square root of the sample variance, s^2, which is itself defined by

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1). \tag{A.11}$$

This can also be calculated as

$$s^2 = \left\{ \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \Big/ n \right\} /(n-1). \tag{A.12}$$

As an example of the calculations, consider the five measurements of the medullary widths, in microns, of cat hairs ($n = 5$) in Table A.3

Then, from (A.10)

$$\bar{x} = \sum_{i=1}^{n} x_i/n = 94.700/5 = 18.9400.$$

Table A.3 Five measurements $x_1, \ldots, x_5$ of medullary widths in microns of cat hairs.

x_1	x_2	x_3	x_4	x_5
17.767	18.633	19.067	19.300	19.933

From (A.12)

$$s^2 = \left\{ \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \Big/ n \right\} / (n-1)$$

$$= (1796.220 - 94.700^2/5)/4 = 0.6505$$

and the sample standard deviation is

$$s = \sqrt{(0.6505)} = 0.8065.$$

Note that the sample mean and standard deviation are quoted to one more decimal place than the original measurements.

Another population parameter is the *quantile*. This is the parameter that specifies the proportion of the population that is below a certain value. Thus the $100p\%$ quantile is the value x of a random variable X such that $\Pr(X \leq x) = p$. This is illustrated in Figure A.1. The median is the special case when $p = 0.5$.

A.3.2 Normal Distribution

When considering data in the form of counts, the variation in the possible outcomes can be represented by a function known as a *probability function*. The variation in measurements, which are continuous, may also be represented mathematically by a function, known as a *probability density function*. Probability functions and probability density functions are both examples of *probability models*.

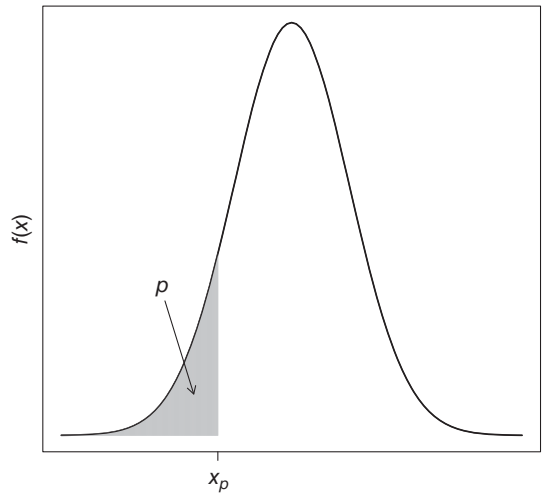

Figure A.1 Quantile x_p of order p of a probability density function $f(x)$.

As an example of a probability model for a continuous measurement, consider the estimation of the quantity of alcohol in blood. From experimental results, it has been determined that there is variation in the measurements, x (in g/kg), provided by a certain procedure. The variation is such that it may be represented by a probability density function that in this case is unimodal, symmetric, and bell-shaped. The particular function that is used here is the *Normal* or *Gaussian* probability density function (named after the German mathematician Carl Friedrich Gauss, 1777–1855).

The binomial distribution required the number of trials and the probability of a success to be known in order that the probability function could

be defined. Two characteristics (or parameters) of the measurement are required to define the Normal probability density function. These are the mean, or expectation, μ, and the standard deviation, σ. The mean may be thought of as a *measure of location* to indicate the size of the measurements. The standard deviation may be thought of as a *measure of dispersion* to indicate the variability in the measurements. The square of the standard deviation, the variance, is denoted σ^2. Given these parameters, the Normal probability density function for x, $f(x \mid \mu, \sigma^2)$, is given by

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \tag{A.13}$$

The function takes its maximum value when $x = \mu$, it is defined on the whole real line for $-\infty < x < \infty$ and is always positive. The area underneath the function is 1, since x has to lie between $-\infty$ and ∞.

As an example of its use, consider blood alcohol levels. In some countries if the alcohol level in blood is estimated to be greater than 0.80 g/kg a person is considered to be under the influence of alcohol. The variability inherent in a measurement, x, of alcohol quantity is known from previous experiments to be such that it is Normally distributed about the true value μ with variance, σ^2, of 0.005. Consider a person whose true unknown quantity, μ, of alcohol in the blood is 0.70 *g/kg*. The probability density function $f(x \mid \mu, \sigma^2)$ for the measurement of the quantity of

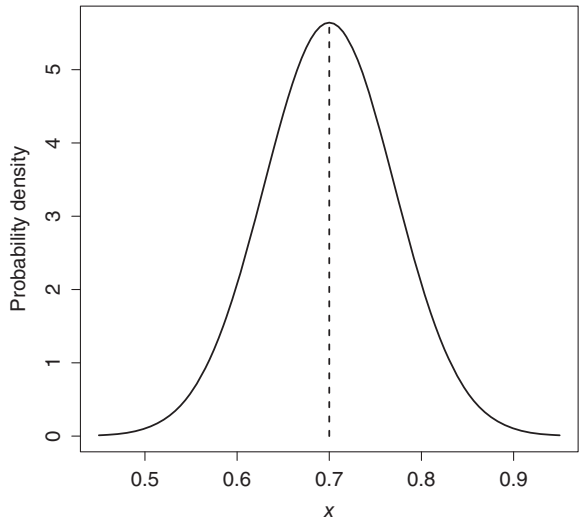

Figure A.2 Probability density function for a Normal distribution, with mean 0.7 and variance 0.005.

alcohol in the blood is then obtained from (A.13) with the substitution of 0.70 for μ and 0.005 for σ^2. The function is illustrated in Figure A.2. Note the labelling of the ordinate as 'probability density'. The reasoning for this is described later in this section. In particular it is possible for the probability density function to take values greater than 1.

There is a special case of zero mean ($\mu = 0$) and unit variance ($\sigma^2 = 1$). The Normal probability density function is then

$$f(z \mid 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \qquad \text{(A.14)}$$

where z is used instead of x to denote the special nature of parameter values of zero mean and unit

variance. The Normal probability density function is so common that it has special notation. If a random variable Z is Normally distributed with mean 0 and variance 1, it is denoted

$$Z \sim N(0, 1),$$

where Z is the random variable corresponding to z and the conditioning on $\mu = 0$ and $\sigma^2 = 1$ on the left-hand side has been omitted for clarity. This distribution is known as a standard Normal distribution. In general, a Normally distributed measurement, X, say, with mean μ and variance σ^2, may be said to be such that

$$(X \mid \mu, \sigma^2) \sim N(\mu, \sigma^2).$$

The first symbol within parentheses conventionally denotes the mean, the second conventionally denotes the variance. It is not always necessary for the notation to make explicit the dependence of X on μ and σ^2. The distributional statement may then be denoted

$$X \sim N(\mu, \sigma^2),$$

and such abbreviated notation is used often.

The determination of probabilities associated with Normally distributed random variables is made possible by a process known as standardisation, whereby a general Normally distributed random variable is transformed into one that has a standard Normal distribution. Let

$$Z = (X - \mu)/\sigma.$$

Then $E(Z) = 0$ and $Var(Z) = 1$ and the random variable Z has a standard Normal distribution. Notice that standardisation requires variability, as represented by σ, to be taken into account. For example, the division by σ ensures the resulting statistic is dimensionless.

Consider the following numerical example of blood alcohol measurements using the parameter values above. Let X be the random variable of measurements of blood alcohol for a particular person, with x denoting the value of a particular measurement. Suppose the true, unknown, level of alcohol in the person's blood is, as before, $\mu = 0.70$ g/kg and the standard deviation in the measurement is σ, the square root of 0.005, which equals 0.07 (to two decimal places). Suppose the measurement x of the blood alcohol quantity recorded by the measuring apparatus is 0.85 g/kg, which is over the permitted limit of 0.80 g/kg. The variance σ^2 is assumed known as it has been estimated from many previous experiments with the measuring apparatus and it is assumed to be a constant, independent of μ. Substitution of $x = 0.85$ g/kg, $\mu = 0.70$ g/kg, though unknown in this case, and $\sigma^2 = 0.005$ into (A.13) gives

$$f(x) = f(0.85) = \frac{1}{\sqrt{0.01\pi}} \times \exp\left(-\frac{(0.85 - 0.70)^2}{0.01}\right) = 0.60, \tag{A.15}$$

see Figure A.2. In practice, what is of interest is the probability that the true blood alcohol level is greater than 0.80 g/kg, when the instrument provides a measurement of 0.85 g/kg. This requires consideration of a prior distribution for μ and is discussed in detail in Section 4.5.

Consider the continuous case in more detail. The function modelling the variation is known as a probability density function, not a probability function as it does not measure probabilities. An intuitive understanding of the terminology can be gained by considering the following analogy. A cylindrical rod, with circular cross-section, has a density that varies along its length according to some function, f, say. Then its weight over any particular part of its length is the integral of this function f over that part. In the same way, with a probability density function, the probability of a random variable lying in a certain interval is the integral of the corresponding density function over the interval. Thus the probability of the measurement of the blood alcohol quantity x lying within a certain interval would be the integral of $f(x)$ over this interval. Note, however, the following theoretical detail. A cross-section of zero thickness of the rod would have zero weight since its volume would be zero. Similarly, the probability of a continuous random variable taking a particular value is zero. In practice, measuring instruments are not sufficiently accurate to measure to an infinite number of decimal places, and this problem does not arise so long as one

determines the probability of a measurement lying within a particular interval and does not attempt to calculate the probability of a measurement taking a particular value. (see Section 3.5.5 for an application of this idea.)

This probability cannot be determined analytically and reference has to be made to appropriate statistical packages.

Let Z be a random variable with a standard Normal distribution, thus

$$Z \sim N(0, 1).$$

There is a special notation to denote the probability, known as the cumulative distribution function, that Z is less than a particular value z. The probability that Z is less than z, $\Pr(Z < z)$, is denoted $\Phi(z)$. Certain values of z are used commonly in the discussion of significance probabilities. See, for example, Section 3.6.1, particularly those values for which $1 - \Phi(z)$ is small, and some of these are tabulated in Table A.4.

The probability $1 - \Phi(z)$ is the probability that $Z > z$.[3] Corresponding probabilities for absolute values of Z may be deduced from the tables by use of the symmetry of the Normal distribution. By symmetry,

$$\begin{aligned}\Phi(-z) &= \Pr(Z < -z) = \Pr(Z > z) = 1 - \Pr(Z < z) \\ &= 1 - \Phi(z).\end{aligned}$$

[3] As the variable Z is continuous there is no need to be concerned with $Z = z$ as, for reasons explained earlier, such an event has probability zero.

Table A.4 Values of cumulative distribution function $\Phi(z)$ and its complement $1 - \Phi(z)$ for the standard Normal distribution for given values of z.

z	$\Phi(z)$	$1 - \Phi(z)$
1.6449	0.950	0.050
1.9600	0.975	0.025
2.3263	0.990	0.010
2.5758	0.995	0.005

Thus

$$\begin{aligned}\Pr(|\,Z\,| < z) &= \Pr(-z < Z < z)\\ &= \Pr(Z < z) - \Pr(Z < -z)\\ &= \Phi(z) - \Phi(-z)\\ &= 2\Phi(z) - 1.\end{aligned}$$

Particular, commonly used, values of z with the corresponding probabilities for the absolute values of z are given in Table A.5.

Figure A.3 illustrates the probabilities for the following events:

- (a) $\Pr(Z > 1) = 0.159$,
- (b) $\Pr(Z > 2) = 0.023$,
- (c) $\Pr(|\,Z\,| < 2) = \Pr(-2 < Z < 2) = 0.954$,
- (d) $\Pr(Z > 2.5) = 0.006$.

An interval, known as a confidence interval, for the mean μ of a random variable X with an

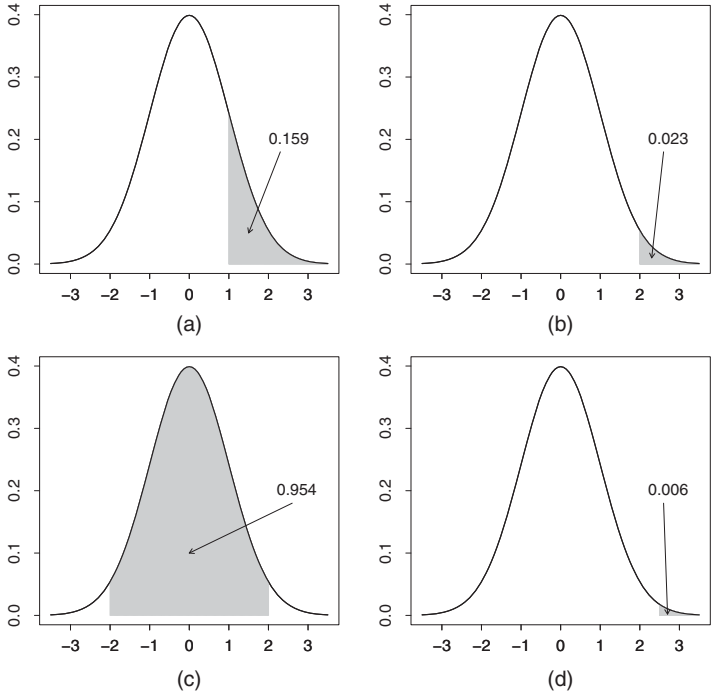

Figure A.3 Selected area probabilities for a standard Normal random variable Z; (a) Pr(Z > 1), (b) Pr(Z > 2), (c) Pr(| Z |< 2), and (d) Pr(Z > 2.5).

$N(\mu, \sigma^2)$ distribution may be determined from the expression

$$\Pr\left(-z_{\alpha/2} < \frac{X-\mu}{\sigma} < z_{\alpha/2}\right) = 1-\alpha, \quad \text{(A.16)}$$

where $\Pr(Z > z_{\alpha/2}) = \Pr(Z < -z_{\alpha/2}) = \alpha/2$ and $Z = (X-\mu)/\sigma \sim N(0,1)$. Rearrangement of (A.16) shows

$$\Pr(X - z_{\alpha/2}\sigma < \mu < X + z_{\alpha/2}\sigma) = 1-\alpha. \quad \text{(A.17)}$$

Replacement of the random variable X with an observation x gives the interval $x - z_{\alpha/2}\sigma < \mu < x + z_{\alpha/2}\sigma$, which is said to be a $100(1-\alpha)\%$ confidence interval for θ. For example, if $\alpha = 0.05$, the 95% confidence interval for θ is

$$(x - 1.96\sigma, x + 1.96\sigma),$$

where the figure 1.96 is taken from Table A.5. Note that a 95% confidence interval for μ means that if an experiment is repeated many times under identical conditions, 95% of the confidence intervals estimated will cover the true value of the parameter of interest. This does not mean that there is a 95% probability that the true value is in the estimated interval. For further details see Kaye (1987b) and Section 4.2.1.

Consider the mean $\bar{X}$ of a sample of size n. Then it can be shown that $\bar{X} \sim N(\mu, \sigma^2/n)$. A similar argument to the one earlier shows that the $100(1-\alpha)\%$ confidence interval for μ, given

Table A.5 Probabilities for absolute values from the standard Normal distribution function.

z	$\Phi(z)$	$\Pr(\lvert Z \rvert < z) = 2\Phi(z) - 1$	$\Pr(\lvert Z \rvert > z)$
1.6449	0.950	0.90	0.10
1.9600	0.975	0.95	0.05
2.3263	0.990	0.98	0.02
2.5758	0.995	0.99	0.01

observations $x_1, \ldots, x_n$, is

$$(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}).$$

Some variables, including blood alcohol level, can only take positive values. If the mean is sufficiently far away from zero, in units of standard deviations, then the probability the variable takes a value less than zero can effectively be discounted.

In some cases, the distribution may be *positively skewed* in the sense that the right-hand-tail of the distribution is much longer than the left-hand-tail and the distribution is asymmetric (e.g. the mean is greater than the median). (A distribution in which the left-hand-tail of the distribution is much longer than the left-hand-tail and the mean is less than the median is said to be *negatively skewed*.) For positively skewed distributions, a transformation to the logarithm of the variable of interest will often produce a variable, which is more symmetric than the original variable and for which inferences based on the Normal distribution may be used. Care must then be taken to remember to transform the results back to the original measurements for the final summary.

A.3.2.1 Normal Approximations to the Binomial and Poisson Distributions

One of the advantages of the Normal distribution is that it can be used as an approximation to other distributions in situations where it may be

impractical or just simply tedious (such as in the absence of a suitable statistical package to do the sums) to use the other distributions. Two examples are the binomial and Poisson distributions. These are two discrete distributions where it is tedious to evaluate exact probabilities for large numbers of events. For example, whilst possible, it is tedious to evaluate exactly the probability of less than 531 heads in 1000 tosses of a fair coin.

Let X be a random variable with a binomial distribution with n trials and success probability θ so that $\mathrm{E}(X) = n\theta$ and $\mathrm{Var}(X) = n\theta(1-\theta)$. For n large and θ not too close to 0 or 1, the distribution of X may be approximated by a Normal distribution with the same mean and variance. Thus, approximately,

$$X \sim N(n\theta, n\theta(1-\theta)).$$

Also, $\mathrm{E}(X/n) = \theta, \mathrm{Var}(X/n) = \theta(1-\theta)/n$ and, again approximately,

$$X/n \sim N(\theta, \theta(1-\theta)/n). \tag{A.18}$$

In answer to the question posed immediately earlier, let X denote the number of heads in 1000 tosses of a fair coin. Then

$$\begin{aligned}&\Pr(X < 531 \mid n = 1000, \theta = 0.5)\\ &= \sum_{x=0}^{530} \binom{1000}{x} 0.5^x \, 0.5^{1000-x},\end{aligned}$$

a sum which can be evaluated, given time.[4] Alternatively, the Normal approximation is the following:

$$\begin{aligned}&\Pr(X < 531 \mid n = 1000, \theta = 0.5)\\ &\quad \simeq \Phi((530.5 - 500)/\sqrt{2}50) = \Phi(1.93)\\ &\quad = 0.9732,\end{aligned}$$

where 0.5 is added to 530 to allow for the approximation of a discrete distribution by a continuous distribution. For the discrete distribution, X takes only integer values $(\ldots, 529, 530, 531, \ldots)$, whereas for the continuous distribution, X can take any value. In this example, the value 530.5 is chosen as being the value midway between 530 and the value 531 immediately above it. The exact probability evaluated using statistical software is 0.9732. The approximation is excellent.

Let $y_1, \ldots, y_n$ be n observations for a Poisson distribution with mean λ. Let $\bar{Y}$ be the random variable corresponding to the sample mean. The expectation of $\bar{Y}$ is $\mathrm{E}(\sum_{i=1}^{n} Y_i/n) = \sum_{i=1}^{n} \mathrm{E}(Y_i)/n = n\lambda/n = \lambda$ and the variance of $\bar{Y}$ is $\mathrm{Var}(\sum_{i=1}^{n} Y_i/n) = \sum_{i=1}^{n}(\mathrm{Var}(Y_i)/n^2 = n\lambda/n^2 = \lambda/n$. For large n, the distribution of $\bar{Y}$ may be approximated by a Normal distribution such that

$$\bar{Y} \sim N(\lambda, \lambda/n). \tag{A.19}$$

[4]The probability that $X < 531$ is the probability that $X \leq 530$ since X is an integer. The statement $X < 531$ implies X takes one of the values $0, \ldots, 530$.

A.3.3 Jeffreys' Prior Distributions

In Bayesian probability, Jeffreys' prior distribution, named after Sir Harold Jeffreys (1891–1989), is a non-informative (objective) prior distribution for a parameter space (Jeffreys, 1983). Examples include

- $f(\mu) \propto 1, -\infty < \mu < \infty$, for the distribution of the mean μ of a Normally distributed random variable $X \sim N(\mu, \sigma^2)$;
- $f(\sigma) \propto 1/\sigma$, $\sigma > 0$, for the distribution of the standard deviation σ of a Normally distributed random variable $X \sim N(\mu, \sigma^2)$;
- $f(\lambda) \propto 1/\sqrt{\lambda}$, $\lambda > 0$, for the distribution of the mean λ of a Poisson distributed random variable $X \sim Po(\lambda)$;
- $f(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, $\theta \in [0, 1]$, for the distribution of the probability θ of a Bernoulli trial that is a 'success' with probability θ and is a 'failure' with probability $(1 - \theta)$.

Note that the prior may not be proper in that it does not integrate to a finite value over its range. However, when combined with a likelihood function, a proper posterior distribution is obtained. See, for example, Lee (2012), Taroni et al. (2010).

A.3.4 Student's *t*-Distribution

In practice, the standard deviation σ of data from a Normal distribution is rarely known

and it is estimated from the data by the sample standard deviation s. Consider n independent, identically distributed Normal random variables $X_1, X_2, \ldots, X_n$ such that

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n.$$

Then the random variable $\bar{X}$ corresponding to the mean of $X_1, \ldots, X_n$, and given by

$$\bar{X} = \sum_{i=1}^{n} X_i/n$$

has itself a Normal distribution, such that

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

The transformed, standardised variable Z, defined as

$$Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$$

has a standard Normal $N(0, 1)$ distribution.

Precision as a statistical concept is the reciprocal of the variance. Thus to double the precision of an estimator of a parameter it is necessary to increase the number of observations by a factor of four.

If the standard deviation, σ, is not known and it is replaced by the sample standard deviation S ($S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$) corresponding to its estimate s, the resulting statistic is

$$(\bar{X} - \theta)/(S/\sqrt{n}). \qquad \text{(A.20)}$$

This statistic does not have a standard Normal distribution. It is the ratio of the functions of

two random variables $\bar{X}$ and S. The distribution is known as a *Student's t-distribution* and the corresponding statistic is known as a t-statistic. ('Student' was the pseudonym of W.S. Gosset, 1876–1937.) The distribution is symmetric about zero. The extra uncertainty induced by replacing σ with an estimate s leads to the t-distribution having greater dispersion than the standard Normal distribution. Also, the distribution depends on the sample size, n. In particular, if the standard deviation s is estimated from a sample of n observations $x_1, \ldots, x_n$ for use in the statistic (A.20) then the value $(n - 1)$ is known as the number of *degrees of freedom* associated with the t-statistic. The degrees of freedom are determined from the denominator of the expression used to derive s. Informally, the number of degrees of freedom may be considered as the number of observations free to estimate s after 1 has been deducted from n to estimate $\bar{x}$. Given the values of $(n - 1)$ observations and the value of the mean $\bar{x}$ on n observations, it is possible to determine the value of the n-th observation from the expression

$$x_n = n\bar{x} - \sum_{i=1}^{n-1} x_i.$$

As n increases, the t-distribution approaches the standard Normal distribution. As with the standard Normal distribution, the associated probabilities cannot be determined analytically and reference has to be made to statistical software. See Section 4.5.2 for an example.

Table A.6 Percentage points $t_{(n-1)}(P)$ for the t-distribution for given values of sample size n, degrees of freedom $(n-1)$ and P, and the corresponding point $z(P)$ for the standard Normal distribution.

$P\%$	$(100-P)\%$	n	$(n-1)$	$t_{(n-1)}(P)$	$z(P)$
95	5	10	9	1.812	1.645
95	5	20	19	1.725	1.645
99	1	10	9	2.764	2.326
99	1	20	19	2.528	2.326
99.5	0.5	10	9	3.250	2.576
99.5	0.5	20	19	2.861	2.576

Some probabilities for the t-distribution are given in Table A.6 where $t_{(n-1)}(P)$ is the value of t from a t-distribution with $(n-1)$ degrees of freedom (denoted $t_{(n-1)}$) such that the probability the random variable T (with a $t_{(n-1)}$ distribution) is greater than $t_{(n-1)}(P)$ is $P/100$. For example, when the sample size n is 20, the probability that T is greater than 2.528 is $1/100$ or 0.01.

There is a more general form of the t-distribution, which is not centred about zero, known as a *non-central t-distribution*. There are three parameters which will be denoted μ, λ, and ν. If X has such a non-central t-distribution then the transformed variable $Y = (X-\mu)/\lambda$ has a (central) t-distribution with ν degrees of freedom. An example of the use of this distribution is given in Section 7.3.1 to determine the value of the

numerator in the evaluation of glass fragments, where μ is a control mean and λ is an estimate of the standard deviation of the refractive index of the population of glass fragments from which the recovered fragments have come. The value of the numerator is the value of the central t-density at the appropriate point, with an adjustment by multiplication of the density value by a factor of $1/\lambda$ to allow for the standardisation.

A.3.5 Gamma and Chi-Squared Distributions

The gamma distribution is a conjugate prior for various types of inverse scale or rate parameters. It is parametrised by a positive shape parameter, here denoted $\alpha > 0$ and a positive rate parameter, here denoted $\beta > 0$. The probability density function for a random variable X is

$$f(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; \;\; x > 0.$$

The expectation is $\mathrm{E}(X) = \alpha/\beta$, the variance is $\mathrm{Var}(X) = \alpha/\beta^2$ and $\Gamma(\alpha)$ is the gamma function where $\Gamma(\alpha) = (\alpha - 1)!$ for integer α and $\Gamma(1/2) = \sqrt{\pi}$.

A special case of the gamma distribution known as the chi-squared distribution occurs when α is denoted as $\nu/2$ and $\beta = 1/2$. The parameter ν is the degrees of freedom associated with the

distribution. The probability density function for a random variable X is

$$f(x \mid \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu-2)/2} e^{-x/2}; \;\; x > 0.$$

The expectation is $\mathrm{E}(X) = \nu$ and the variance is 2ν. Examples of the chi-squared distribution with 1, 2, 5, 10, 20, and 50 degrees of freedom are shown in Figure A.4. The statistic $S^2(n-1)/\sigma^2$ of n normally distributed random variables has a chi-squared distribution with $(n-1)$ degrees of freedom.

A.3.6 Inverse Gamma and Inverse Chi-Squared Distributions

The inverse gamma distribution is the distribution of the reciprocal of a variable with a gamma distribution. It arises as a marginal distribution for the variance of a Normal distribution if an uninformative prior is used and as an analytically tractable conjugate prior if an informative prior is required. The probability density function for a random variable X with a shape parameter $\alpha(> 0)$ and scale parameter $\beta(> 0)$ is

$$f(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}; \;\; x > 0.$$

The expectation is $\mathrm{E}(X) = \beta/(\alpha - 1), \;\; (\alpha > 1)$ and the variance is $\mathrm{Var}(X) = \beta^2/\{(\alpha-1)^2(\alpha-2)\}, \;\; (\alpha > 2)$.

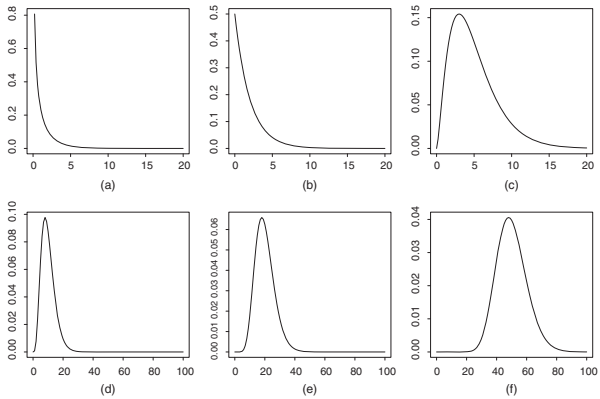

Figure A.4 Probability density function for chi-squared distributions with k degrees of freedom and (a) $k = 1$, (b) $k = 2$, (*c*)k = 5, (d) $k = 10$, (*e*)k = 20, and (f) $k = 50$.

As with the gamma distribution, there is a special case of the inverse gamma distribution known as the inverse chi-squared distribution that occurs when α is written as $\nu/2$ and $\beta = 1/2$. The parameter ν is the degrees of freedom associated with the distribution. The probability density function for a random variable X with an inverse chi-squared distribution with ν degrees of freedom is

$$f(x \mid \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{-(\nu+2)/2} e^{-1/2x}; \quad x > 0.$$

The expectation is $\mathrm{E}(X) = 1/(\nu - 2)\ (\nu > 2)$ and the variance is $\mathrm{Var}(X) = 2/\{(\nu - 2)^2(\nu - 4)\}$ $(\nu > 4)$.

A.3.7 Beta Distribution

Consider an example in which it is desired to know the proportion of a consignment that is illicit drugs. This example has been discussed in Section 4.2 with reference to a consignment of tablets. The number of tablets that are illicit is R and the consignment size is N. The proportion of illicit tablets is then R/N, which takes a finite number of values, depending on the value of R, ranging from $0/N$ to N/N in steps of $1/N$. As N increases, this proportion becomes closer to a continuous measurement, over the interval $(0, 1)$. The uncertainty in a continuous random variable that is a proportion can be modelled by the Beta distribution. Denote the random variable by θ. For a consignment of drugs, assume that it is representative of a

super-population of drugs in which the proportion of illicit tablets is θ $(0 < \theta < 1)$. See Smith and Charrow (1975) and Finney (1977) for comments about super-populations and also Section 2.5.5. For example, the consignment may be known to have come from a particular location and θ is the proportion of units in the super-population that contain drugs. In order to make probability statements about θ, it is necessary to have a probability distribution for θ to represent the uncertainty in θ. This uncertainty may simply be uncertainty in one's knowledge of the exact value of θ, uncertainty that may arise because the consignment is considered as a random sample from a super-population. The Bayesian philosophy permits this uncertainty to be represented as a probability distribution. The beta distribution is the most common distribution for θ, characterised by two parameters, denoted here as α and β with probability density function

$$f(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq \theta \leq 1, \quad \text{(A.21)}$$

denoted $Be(\alpha, \beta)$, where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \text{(A.22)}$$

and Γ is the gamma function of Section A.3.5. The expectation is $\mathrm{E}(X) = \alpha/(\alpha+\beta)$ and the variance is $\mathrm{Var}(X) = \alpha\beta/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$. The function $B(\alpha, \beta)$ is known as the *beta function*.

Note the special case

$$B(m+1,1) = \frac{\Gamma(m+1)\Gamma(1)}{\Gamma(m+2)} = \frac{1}{m+1}.$$

Examples of the beta distribution with parameters $(\alpha, \beta) = (5,5), (3,3), (1,1), (5,1), (2,5),$ and $(0.5, 0.5)$ are shown in Figure A.5.

The use of the beta distribution in this context is described in Aitken (1999). Values for α and β may be chosen subjectively to represent the scientist's prior beliefs before inspection about the proportion of the units in the consignment (as a random sample from the super-population) that contain drugs. A large value of α relative to β would imply a belief that θ was high. Larger values of α and β would correspond to higher certainty about the value of θ. A detailed discussion is given in Aitken (1999) and summarised in Section 4.3.1. In many cases, the scientist will not wish to quantify their prior beliefs and will wish to remain neutral. This can be done by choosing $\alpha = \beta = 1$. Also, as shown in Aitken (1999) for variations in α and β, when both are small, the evidence from the sample will soon reduce the effect of the values of α and β considerably. This is intuitively reasonable: little prior information is soon subsumed by the data.

The beta distribution on $[0, 1]$ can be generalised to the interval $[a, b]$ with $a < b$. The density function is

$$f(x \mid \alpha, \beta, a, b) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}B(\alpha,\beta)};$$
$$a < x < b,\ \alpha, \beta > 0. \qquad \text{(A.23)}$$

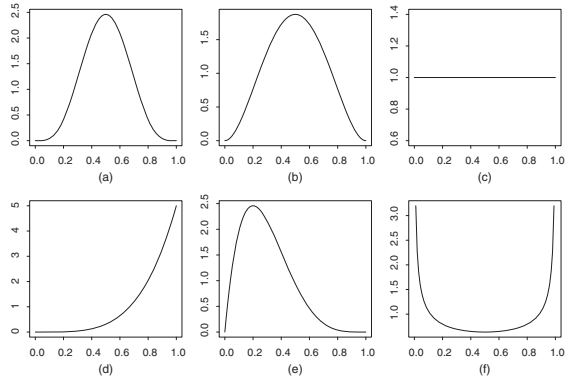

Figure A.5 Probability density functions for beta distributions with parameters (α, β) and (a) $(\alpha, \beta) = (5, 5)$, (b) $(\alpha, \beta) = (3, 3)$, (c) $(\alpha, \beta) = (1, 1)$, (d) $(\alpha, \beta) = (5, 1)$, (e) $(\alpha, \beta) = (2, 5)$, and (f) $(\alpha, \beta) = (0.5, 0.5)$.

The expectation and variance are

$$\mathrm{E}(X) = \frac{\alpha b + \beta a}{\alpha + \beta},$$

$$\mathrm{Var}(X) = \frac{\alpha\beta(b-a)^2}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The special case when $a = 0$ and $b = n\ (> 0)$ has density function

$$f(\theta \mid \alpha, \beta, n) = \frac{1}{B(\alpha,\beta)} \frac{\theta^{\alpha-1}(n-\theta)^{\beta-1}}{n^{\alpha+\beta-1}}; \quad 0 < \theta < n. \tag{A.24}$$

A.3.8 Dirichlet Distribution

The example in which it is desired to know the proportion of a consignment that is illicit drugs may be generalised to a situation in which there may be several drug types (say k) and it is desired to know the proportions in each type. Consider again a consignment of tablets of size N. Denote the number of tablets in each of the k types as $R_i, i = 1, \ldots, k$. The proportion of tablets of type i is then R_i/N, which takes a finite number of values, depending on the value of R_i, ranging from $0/N$ to N/N in steps of $1/N$. As N increases, these proportions become closer to a continuous measurement, over the interval $(0, 1)$. The uncertainty in a set of random variables that are proportions and for which the sum is 1 is modelled by a generalisation of the beta distribution, known as the Dirichlet distribution, named after the German mathematician, P.G.L. Dirichlet

(1805–1859). This generalisation is analogous to the generalisation by the multinomial distribution (Section A.2.4) of the binomial distribution (Section A.2.3).

Denote the set of random variables by $\theta_i, i = 1, \ldots, k$, which are such that $\sum_{i=1}^{k} \theta_i = 1$. The beta distribution is the case described here for which $k = 2$ and, conventionally, θ_1 is denoted θ and $\theta_2 = 1 - \theta_1 = 1 - \theta$. For a consignment of drugs, assume, as before, that it is representative of a super-population of drugs in which the proportion of tablets in each of the k categories is $\theta_i, i = 1, \ldots, k; 0 < \theta_i < 1; \sum_{i=1}^{k} \theta_i = 1)$. For example, the consignment may be thought to have come from a particular location and the set $\{\theta_i; i = 1, \ldots, k\}$ are the proportions of units in the super-population, which fall into the k categories. In order to make probability statements about $\{\theta_i\}$, it is necessary to have a probability distribution for $\{\theta_i\}$ to represent the uncertainty in $\{\theta_i\}$. This uncertainty may simply be uncertainty in one's knowledge of the exact values of $\{\theta_i\}$, uncertainty which may arise because the consignment is considered as a random sample from a super-population. The Dirichlet distribution is the most common distribution for $\{\theta_i; i = 1, \ldots, k\}$ with probability density function

$$f(\theta_1, \ldots, \theta_k \mid \alpha_1, \ldots, \alpha_k) = \frac{\theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}}{B(\alpha_1, \ldots, \alpha_k)},$$

$$0 < \theta_i < 1, i = 1, \ldots, k, \ \sum_{i=1}^{k} \theta_i = 1, \quad \text{(A.25)}$$

where

$$B(\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \cdots + \alpha_k)}.$$

The mean $\mathrm{E}(\theta_i)$ of θ_i is $\alpha_i/(\sum_{i=1}^{k} \alpha_i)$ and the variance $\mathrm{Var}(\theta_i)$ of θ_i is $\mathrm{E}(\theta_i)(1 - \mathrm{E}(\theta_i))/(1 + \sum_{i=1}^{k} \alpha_i)$. The $\{\theta_i\}$ add to 1 so they are correlated. The covariance $\mathrm{Cov}(\theta_i, \theta_j)$, $i \neq j$, between θ_i and θ_j is given by $\mathrm{Cov}(\theta_i, \theta_j) = -\mathrm{E}(\theta_i)\mathrm{E}(\theta_j)/(1 + \sum_{i=1}^{k} \alpha_i)$. Note that this is negative; given the value of θ_i, the range of values for θ_j is reduced from $(0,1)$ to $(0, 1 - \theta_i)$.

The Dirichlet distribution is characterised by k parameters, $\{\alpha_1, \ldots, \alpha_k\}(\alpha_i > 0; i = 1, \ldots, k)$. Values for $\{\alpha_1, \ldots, \alpha_k\}$ may be chosen subjectively to represent the scientist's prior beliefs before inspection about the proportions of the units in the consignment (as a random sample from the super-population) for each of the k categories.

Consider a single-locus marker for DNA profiling. Let (X_1, X_2) be the sample frequencies of the two alleles of the locus found on a crime scene profile. The overall sample size, from which X_1 and X_2 are obtained is n. Let $X_3 = n - X_1 - X_2$, the sample frequency of the combination of all other alleles at that locus. The corresponding population relative frequencies are θ_1, θ_2, and θ_3 with $\sum_{i=1}^{3} \theta_i = 1$. The Dirichlet distribution provides a convenient prior distribution for the $\{\theta_i\}$, with three categories, $k = 3$.

Further details of the previous example and inferences that may be drawn from study of the

crime scene profile are given in Balding (1995). An example of the use of the Dirichlet distribution as a prior for a multinomial likelihood for blood grouping data is given in Leonard and Hsu (1999, pp. 195–196). Applications to forensic match probabilities are described in Lange (1995) and to sample size estimation with categorical responses in Aitken and Mavridis (2009); see also Section 7.2.3.

A.3.9 Multivariate Normal Distribution and Correlation

Often, more than one characteristic is of interest, e.g. the refractive index, the density and various elemental compositions for window glass. The data (measurements of these characteristics) are referred to as *multivariate data* and in the special case where only two characteristics are measured they are known as *bivariate data*. Let the measurements be denoted as a vector $\mathbf{x}$.[5] There is a notational convention as to how a vector is written. A vector $\mathbf{x}$ is conventionally written in bold script and expanded as a column (illustrated here with p variables)

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$

[5] A *vector* in mathematics is a list of more than one characteristic associated with a unit of interest. With only one characteristic, the corresponding term is *scalar*.

The transpose of a column vector to a row vector is indicated with a superscript T such that $\mathbf{x}^T = (x_1, x_2, \ldots, x_p)$. For bivariate data $p = 2$. In the example of window glass, x_1 would be the value of the refractive index, x_2, the value of the density and the elemental compositions would be denoted $x_3, x_4, \ldots, x_p$. For continuous data, the vector $\mathbf{x}$ has a probability density function, just as the individual characteristics have.

If the characteristics are independent (Section 1.7.8) then the joint probability density function $f(\mathbf{x})$ is the product of the individual probability density functions. Thus

$$f(\mathbf{x}^T) = f(x_1, \ldots, x_p) = \prod_{i=1}^{p} f(x_i), \qquad \text{(A.26)}$$

which may be thought of as an extension of the third law of probability for independent events (1.10).

If the characteristics are not independent, however, such an approach is not possible. Assume the measurements of these characteristics are Normally distributed and dependent. The measurements are said to be *correlated*. A multivariate analogue of the Normal distribution may be obtained. The multivariate mean $\boldsymbol{\mu}$ is the vector formed by the means of the individual variables. Instead of a variance σ^2 there is a square $(p \times p)$ symmetric matrix Σ of variances

and *covariances*. Some properties of matrices are given in Appendix B.

The matrix Σ is known as the *covariance matrix*. Covariance is a measure of the association between a pair of characteristics and is the product of the individual standard deviations and a factor that measures the *correlation* (degree of linear association) between the two characteristics. The variances of the p variables are located on the diagonal of Σ. The covariances are the off-diagonal terms so that the (i, j)-th cell of Σ contains the covariance between X_i and X_j (the covariance of X_i and X_i is simply the variance of X_i). The correlation between two variables is a parameter, conventionally denoted ρ, which measures the amount of linear association between the variables. It takes values between -1 and 1. Two variables that have a perfect linear relationship with a positive slope (as one increases so does the other) have a correlation of 1 ($\rho = 1$). Two variables which have a perfect linear relationship with a negative slope (as one increases, the other decreases) have a correlation of $\rho = -1$. A correlation of 0 implies that there is no linear association between the two variables. Notice that this does not mean there is no association between the variables, just that there is no linear association.

Denote the variance of X_i by σ_i^2, $(i = 1, \ldots, p)$. The correlation between X_i and X_j is denoted by the correlation coefficient ρ_{ij} $(i = 1, \ldots, p;$

$j = 1, \ldots, p;\ i \neq j$), with $\rho_{ij} = \rho_{ji}$. Then the covariance between X_i and X_j, denoted $\mathrm{Cov}(X_i, X_j)$, is given by

$$\mathrm{Cov}(X_i, X_j) = \rho_{ij}\sigma_i\sigma_j, \quad i = 1, \ldots, p;\ j = 1, \ldots, p.$$
$$\mathrm{Cov}(X_i, X_i) = \rho_{ii}\sigma_i^2 = \sigma_i^2 \text{ since } \rho_{ii} = 1.$$

Denote the determinant of Σ by $|\,\Sigma\,|$ (Section A.1.4) and the inverse by Σ^{-1} (Section B.1.6). Then the probability density function of $\mathbf{X}$ is given by

$$f(\mathbf{x}) = (2\pi)^{-\frac{1}{2}p}\ |\,\Sigma\,|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}. \qquad \text{(A.27)}$$

This may be written in short-hand, equivalent to the univariate case, as

$$(\mathbf{X} \mid \boldsymbol{\mu}, \Sigma) \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma). \qquad \text{(A.28)}$$

Consider the special case of $p = 2$. The multivariate Normal distribution is then called the bivariate Normal distribution. The vector parameters may be written out in full;

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$
$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \qquad \text{(A.29)}$$
$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \sigma_1^{-2} & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & \sigma_2^{-2} \end{pmatrix},$$
$$|\,\Sigma\,|^{\frac{1}{2}} = \sigma_1\sigma_2\sqrt{(1-\rho^2)}.$$

Notice that

$$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \left\{ \frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right\} /(1-\rho^2).$$

The bivariate Normal density function may then be written as

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \times \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x_1-\mu_1)^2}{\sigma_1^2} -2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right\} \right].$$

For the special case in which $\mu_1 = \mu_2 = 0$,

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \times \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2} \right) \right\}.$$

Another special case is when $\rho = 0$. The bivariate Normal density function is then

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2}$$

$$\times \exp\left[-\frac{1}{2}\left\{\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\}\right]$$

which may be written as

$$\frac{1}{2\sqrt{\pi\sigma_1^2}}\exp\left[-\frac{1}{2}\left\{\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right\}\right]$$
$$\times\frac{1}{2\sqrt{\pi\sigma_2^2}}\exp\left[-\frac{1}{2}\left\{\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\}\right].$$

This is the product of the probability density functions for two Normal distributions, one with mean μ_1 and variance σ_1^2 and one with mean μ_2 and variance σ_2^2.

Applications are given in Sections 3.6.3 and 7.6.

A.3.10 Wishart Distribution

The multivariate analogue of the gamma distribution (Section A.3.5) is the Wishart distribution, named after the Scottish mathematician and agricultural statistician, John Wishart (1898–1956). The distribution is the conjugate prior of the inverse covariance matrix of a multivariate random Normal variable. Consider a $p \times p$ positive definite symmetric matrix V. The parameters are Ω, a $p \times p$ positive definite symmetric matrix, and the degrees of freedom $n(> (p-1))$.

The probability density function for V is

$$f(V \mid \Omega, n) = \frac{\mid V \mid^{(n-p-1)/2}}{2^{np/2} \mid \Omega \mid^{n/2} \Gamma_p(n/2)} \times \exp\{-\mathrm{tr}(\Omega^{-1}V)/2\};$$

where 'tr' is the trace of the matrix, the sum of terms on the leading diagonal, and $\Gamma_p(\cdot)$ is the multivariate gamma function

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left(\frac{n}{2} - \frac{j-1}{2}\right).$$

This may be written in short-hand, equivalent to the univariate case, as

$$(V \mid \Omega, n) \sim W(\Omega, n).$$

A.3.11 Inverse Wishart Distribution

The inverse Wishart distribution is the multivariate generalisation of the univariate inverse gamma distribution (Section A.3.6). Let U be a $p \times p$ positive definite matrix following an inverse Wishart distribution with positive definite scale matrix Σ, and n degrees of freedom. Then for $2p < n$, the probability density of U is

$$f(U \mid \Sigma, n) = \frac{\mid \Sigma \mid^{(n-p-1)/2}}{c \mid U \mid^{n/2}} \exp\{-\mathrm{tr}(U^{-1}\Sigma)/2\},$$

where the constant c is given by

$$c = 2^{(n-p-1)p/2}\pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma\left(\frac{n-p-j}{2}\right).$$

This may be written in short-hand, equivalent to the univariate case, as

$$(U \mid \Sigma, n) \sim W^{-1}(\Sigma, n).$$

The mean of an inverse Wishart distribution is given by

$$\mathrm{E}(U) = \frac{\Sigma}{n-2p-2}, \qquad n-2p > 2.$$

It is the distribution of the inverse of a random matrix following a Wishart distribution. Note that if $V \sim W(\Omega, n)$, if follows that $V^{-1} \sim W^{-1}(\Omega^{-1}, n+p+1)$.

An inverse Wishart distribution can be used to model the uncertainty about the covariance matrix of a multivariate random variable. It is the conjugate prior for the covariance matrix of a multivariate Normal distribution. It has been proposed by Bozza et al. (2008) to model the within-source covariance matrix characterising a writer in the context of handwriting evidence (see Section 7.6.2.4).

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

B

Matrix Properties

B.1 MATRIX TERMINOLOGY

A brief introduction to matrices is given here. If further details are desired, good references are Lütkepohl (1996), Mardia et al. (1979), and Press (1982).

A matrix A is a rectangular array of numbers. If A has r rows and c columns it is said to be of *order* $r \times c$ (read as r-by-c). For example, r measurements on c characteristics (or variables) may be arranged in this way. The subscripts for A denote the cell in which the item is located. Thus a_{ij} is the member of the (i, j)-th cell, the cell in row i and column j of the matrix. A matrix A is sometimes denoted $\{a_{ij}\}$.

An example is the matrix of variances and covariances Σ, introduced in Section A.3.9 and known as the *covariance matrix*. If $r = c$, the matrix is said to be square of order r, and an example of a square matrix is the covariance matrix. The diagonal terms are the variances and the off-diagonal terms (those in the top-right-hand and bottom-left-hand corners of 2×2 matrices)

are the covariances. The covariances are equal since the covariance between variables 1 and 2 is the same as the covariance between variables 2 and 1.

B.1.1 The Trace of a Square Matrix

The trace of a square matrix is the sum of the terms in the leading diagonal, that is, the diagonal that runs from the top-left-hand corner to the bottom-right-hand corner of the matrix. Thus, the trace of a 2×2 matrix A is $a_{11} + a_{22}$. This is denoted $\mathrm{tr}(A)$. It only exists if the matrix is square.

B.1.2 The Transpose of a Matrix

The *transpose* of an $r \times c$ matrix $C = \{c_{ij}\}, (i = 1, \ldots, r; j = 1, \ldots, c)$ is a $c \times r$ matrix $D = \{d_{ji}\}$, $(j = 1, \ldots, c; i = 1, \ldots, r)$ such that $d_{ji} = c_{ij}$. The element in the j-th row and i-th column of D is the element in the i-th row and j-th column of C. The transpose of C is denoted C^T.

Let C be a 3×2 matrix

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{pmatrix}.$$

Then

$$C^T = \begin{pmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \end{pmatrix}.$$

A matrix C is symmetric if it is equal to its transpose, $C = C^T$. Such a matrix is of necessity square with $r = c$. An $r \times 1$ matrix is a column vector and its transpose is a $1 \times r$ row vector. For $r = 3$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

and

$$\mathbf{x}^T = (x_1, \ x_2, \ x_3).$$

B.1.3 Addition of Two Matrices

Two matrices of the same order may be added together to give a third matrix of the same order. Let A and B be 2×2 matrices

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Then the matrix $A + B$ is obtained by adding corresponding cell entries together.

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}.$$

Note that $A + B = B + A$.

B.1.4 Determinant of a Matrix

The determinant of a square matrix A is denoted $| A |$. For the 2×2 matrix A in Section (B.1.3), the

determinant is the difference between the product of the leading diagonal terms and the product of the off-diagonal terms. Thus

$$| A |= (a_{11}a_{22} - a_{12}a_{21}).$$

Care has to be taken in understanding the notation. For a matrix, the symbols $| \cdot |$ denote the determinant. For a real number, the symbols $| \cdot |$ denote the positive value of the number. Multiplication of a matrix by a constant, c, say, results in a matrix in which every cell is multiplied by c. Thus

$$cA = \begin{pmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{pmatrix}.$$

and $| cA | = c^2 | A |$. In general, for a $p \times p$ matrix, $| cA | = c^p | A |$.

B.1.5 Matrix Multiplication

Examples are given here of how matrices and row and column vectors may be multiplied together for two-by-two matrices, and row and column vectors with two components. First, consider the multiplication of a vector and a matrix. Let a column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

and the matrix A be as in Section B.1.3. The order of multiplication of a pair of matrices is important

and the number of columns in the first member of the pair must equal the number of rows in the second member. The outcome is a matrix in which the number of rows equals the number of rows in the first member of the pair of matrices and the number of columns equals the number of columns in the second member of the pair of matrices. Thus, multiplication of an $r \times c$ matrix and a $c \times p$ matrix results in an $r \times p$ matrix.

Multiplication of the (1×2) row vector $\mathbf{x}^T$ and the (2×2) matrix A, written as $\mathbf{x}^T A$ gives a (1×2) row vector. Note that a row or column vector may be thought of as a matrix, with only one row or one column. The row vector $\mathbf{x}^T A$ is

$$(x_1 a_{11} + x_2 a_{21}, x_1 a_{12} + x_2 a_{22}).$$

The members of the row vector multiply the corresponding members of the columns of A and the resultant products are then summed.

Multiplication of the (2×2) matrix A and the (1×2) column vector $\mathbf{x}$ written as $A\mathbf{x}$ gives a (2×1) column vector. The column vector $A\mathbf{x}$ is

$$A\mathbf{x} = \begin{pmatrix} a_{11} x_1 + a_{12} x_2 \\ a_{21} x_1 + a_{22} x_2 \end{pmatrix},$$

This is not equal to $\mathbf{x}^T A$. The expression $\mathbf{x}^T A\mathbf{x}$ is

$$(x_1 a_{11} + x_2 a_{21}, x_1 a_{12} + x_2 a_{22}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

which is equal to

$$x_1^2 a_{11} + x_1 x_2 a_{21} + x_1 x_2 a_{12} + x_2^2 a_{22} = x_1^2 a_{11} + x_1 x_2 (a_{21} + a_{12}) + x_2^2 a_{22}.$$

This is simply a number, not a matrix. A 2×2 symmetric matrix has $a_{12} = a_{21}$. Thus, for such a matrix

$$\mathbf{x}^T A \mathbf{x} = x_1^2 a_{11} + 2 x_1 x_2 a_{12} + x_2^2 a_{22}.$$

For the multiplication of two matrices, the rows of the first matrix and the columns of the second are multiplied together component by component. Thus

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix},$$

$$BA = \begin{pmatrix} a_{11}b_{11} + a_{21}b_{12} & a_{12}b_{11} + a_{22}b_{12} \\ a_{11}b_{21} + a_{12}b_{22} & a_{12}b_{21} + a_{22}b_{22} \end{pmatrix}. \quad \text{(B.1)}$$

In general, $AB \neq BA$. However if A and B are symmetric then $a_{12} = a_{21}$, $b_{12} = b_{21}$ and $AB = BA$. For the product AB, A is said to pre-multiply B and B is said to *post-multiply* A.

B.1.6 The Inverse of a Matrix

The square matrix I defined as

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is known as the *identity matrix*. This is because pre- or post- multiplication of another square matrix

A by I leaves A unchanged: $AI = IA = A$. This may be checked from Section B.1.5. The existence of an identity matrix then leads naturally to the concept of an inverse of a matrix. The inverse of a square matrix A is defined as that matrix, denoted A^{-1}, which when used to pre- or post-multiply A gives as a product the identity matrix I. Thus $AA^{-1} = A^{-1}A = I$. For A, as in Section B.1.3, the inverse of A is given by

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \qquad \text{(B.2)}$$

Matrix multiplication AA^{-1} and $A^{-1}A$ verifies that the products are I and that the matrix given in (B.2) satisfies the definition of an inverse.

Note that an inverse only exists if the determinant $|A|$ is non-zero. A matrix for which the determinant is zero is said to be *singular* and an inverse does not exist. This is the matrix equivalent of the non-existence of the reciprocal of the number 0. The two rows and the two columns of a singular 2×2 matrix are equal or proportional. Note, also, that it makes no sense to consider division by a matrix. The operation with matrices that is equivalent to division by a number is multiplication by an inverse of a matrix.

B.1.7 Completion of the Square

A fundamental result for the derivation of probability distributions for hierarchical models in Chapter 7 is that of completion of the square. Integrals evaluated in the derivation of these

models are those of the multivariate Normal density function. The algebra required for their solution is that of the completion of squares for multivariate data.

Consider vectors $\boldsymbol{\theta}$, $\mathbf{a}$ and $\mathbf{b}$, and matrices A and B of appropriate order for the multiplications. Then the general result is that

$$\begin{aligned}(\boldsymbol{\theta}-\mathbf{a})^T A(\boldsymbol{\theta}-\mathbf{a}) + (\boldsymbol{\theta}-\mathbf{b})^T B(\boldsymbol{\theta}-\mathbf{a}) \\ = (\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T(A+B)(\boldsymbol{\theta}-\boldsymbol{\theta}^*) \\ + (\mathbf{a}-\mathbf{b})^T \times (A^{-1}+B^{-1})^{-1}(\mathbf{a}-\mathbf{b}),\end{aligned}$$

where

$$\boldsymbol{\theta}^* = (A+B)^{-1}(A\mathbf{a}+B\mathbf{b}).$$

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

References

Adam, C. (2016). *Forensic Evidence in Court*. Chichester: Wiley.

Adams, B.J. (2003). The diversity of adult dental patterns in the United States and the applications for personal identification. *Journal of Forensic Sciences* 48: 497–503.

Aitchison, J. and Dunsmore, I.R. (1975). *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.

Aitchison, J., Habbema, J.D.F., and Kay, J.W. (1977). A critical comparison of two methods of statistical discrimination. *Applied Statistics* 26: 15–25.

Aitken, C.G.G. (1986). Statistical discriminant analysis in forensic science. *Journal of Forensic Science Society* 26: 237–247.

Aitken, C.G.G. (1991). Report on the international conference on forensic statistics. *Journal of the Royal Statistical Society, Series A* 154: 45–48. Selected papers included on pp. 49–130.

Aitken, C.G.G. (1993). Statistics and the law: report of a discussion session at the Royal Statistical Society Conference, Sheffield, September 1992. *Journal of the Royal Statistical Society, Series A* 156: 301–304.

Aitken, C.G.G. (1999). Sampling – how big a sample? *Journal of Forensic Sciences* 44: 750–760.

Aitken, C.G.G. (2000). Interpretation of evidence and sample size determination. In: *Statistical Science in*

the Courtroom (ed. J.L. Gastwirth), 1–24. New York: Springer-Verlag.

Aitken, C.G.G. (2003). Conviction by probability. *New Law Journal* 153: 1153–1154.

Aitken, C.G.G. (2012). An introduction to a debate. *Law, Probability and Risk* 11: 255–258.

Aitken, C.G.G. (2018). Bayesian hierarchical random effects models in forensic science. *Frontiers in Genetics, Statistical Genetics and Methodology* 7: 1–14.

Aitken, C.G.G., Bring, J., Leonard, T., and Papasouliotis, O. (1997). Estimation of quantities of drugs handled and the burden of proof. *Journal of the Royal Statistical Society, Series A* 160: 333–350.

Aitken, C.G.G., Connolly, T., Gammerman, A., Zhang, G., Bailey, D., Gordon, R., and Oldfield, R. (1996a). Statistical modelling in specific case analysis. *Science & Justice* 36: 245–255.

Aitken, C.G.G., Gammerman, A., Zhang, G., Connolly, T., Bailey, D., Gordon, R., and Oldfield, R. (1996b). Bayesian belief networks with an application in specific case analysis. In: *Computational Learning and Probabilistic Reasoning* (ed. A. Gammerman), 169–184. Chichester: Wiley.

Aitken, C.G.G. and Gammerman, A. (1989). Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society* 29: 303–316.

Aitken, C.G.G. and Gold, E. (2013). Evidence evaluation for discrete data. *Forensic Science International* 230: 147–155.

Aitken, C.G.G. and Huang, C. (2017). Evidence evaluation for hierarchical, longitudinal, binary data using a distance measure. *Statistica Applicata – Italian Journal of Applied Statistics* 27: 213–223.

Aitken, C.G.G. and Lucy, D. (2002). Estimation of the quantity of a drug in a consignment from measurements on a sample. *Journal of Forensic Sciences* 47: 968–975.

Aitken, C.G.G. and Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics* 53: 109–122, with corrigendum 665–666.

Aitken, C.G.G., Lucy, D., Zadora, G., and Curran, J.M. (2006). Evaluation of transfer evidence for three-level multivariate data with the use of graphical models. *Computational Statistics and Data Analysis* 50: 2571–2588.

Aitken, C.G.G. and MacDonald, D.G. (1979). An application of discrete kernel methods to forensic odontology. *Applied Statistics* 28: 55–61.

Aitken, C.G.G. and Mavridis, D. (2009). Sample size determination for categorical responses. *Journal of Forensic Sciences* 54: 135–151.

Aitken, C.G.G. and Nordgaard, A. (2017). The roles of participants' differing background information in the evaluation of evidence; letter to the Editor. *Journal of Forensic Sciences* 63: 648–649.

Aitken, C.G.G., Roberts, P., and Jackson, G. (2010). *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings* (Practitioner Guide No. 1). Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, London: Royal Statistical Society.

Aitken, C.G.G. and Robertson, J. (1987). A contribution to the discussion of probabilities and human hair comparisons. *Journal of Forensic Sciences* 32: 684–689.

Aitken, C.G.G. and Stoney, D.A. (eds.) (1991). *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Aitken, C.G.G. and Taroni, F. (1997). A contribution to the discussion on 'Bayesian analysis of the deoxyribonucleic acid profiling data in forensic identification applications', Foreman et al. *Journal of the Royal Statistical Society, Series A* 160: 463.

Aitken, C.G.G. and Taroni, F. (1998). A verbal scale for the interpretation of evidence; letter to the Editor. *Science & Justice* 38: 279–281.

Aitken, C.G.G., Taroni, F., and Garbolino, P. (2003). A graphical model for the evaluation of cross-transfer

evidence in DNA profiles. *Theoretical Population Biology* 63: 179–190.

Aitken, C.G.G., Wilson, A., Sleeman, R., Morgan, B., and Huish, J. (2017). Distribution of cocaine on banknotes in general circulation in England and Wales. *Forensic Science International* 270: 261–266.

Alberink, I., Sprong, A., Bolck, A., and Curran, J.M. (2014). Quantifying uncertainty in estimations of the total weight of drugs in groups of complex matrices. *Journal of Forensic Sciences* 59: 1614–1621.

Alberink, I., Sprong, A., Bolck, A., and Vergeer, P. (2017). Quantifying uncertainty in estimations of the total weight of drugs of complex matrices: using the Welch-Satterthwaite equation. *Journal of Forensic Sciences* 62: 1007–1014. https://doi.org/10.1111/1556-4029.13351.

Ali, T., Spreeuwers, L., Veldhuis, R., and Meuwly, D. (2015). Sampling variability in forensic likelihood-ratio computation: a simulation study. *Science & Justice* 55: 499–508.

Alladio, E., Martyna, A., Salomone, A., Pirro, V., Vincenti, M., and Zadora, G. (2017). Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes. *Forensic Science International* 271: 13–22.

Alladio, E., Omedei, M., Cisana, S., D'Amico, G., Caneparo, D., Vincenti, M., and Garofano, P. (2018). DNA mixtures interpretation – a proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. *Forensic Science International: Genetics* 37: 143–150.

Allen, R.J., Balding, D.J., Donnelly, P., Friedman, R., Kaye, D.H., LaRue, H., Park, R.C., Robertson, B., and Stein, A. (1995). Probability and proof in *State v. Skipper*: an internet exchange. *Jurimetrics Journal* 35: 277–310.

Allen, T.J., Cox, A.R., Barton, S., Messam, P., and Lambert, J.A. (1998a). The transfer of glass – Part 4: the transfer of glass fragments from the surface of an item to the person carrying it. *Forensic Science International* 93: 201–208.

Allen, T.J., Hoefler, K., and Rose, S.J. (1998b). The transfer of glass – Part 2: a study of the transfer of glass to a person by various methods. *Forensic Science International* 93: 175–193.

Allen, T.J., Hoefler, K., and Rose, S.J. (1998c). The transfer of glass – Part 3: the transfer of glass from a contaminated person to another uncontaminated person during a ride in a car. *Forensic Science International* 93: 195–200.

Allen, T.J., Locke, J., and Scranage, J.K. (1998d). Breaking of flat glass – Part 5: size and distribution of fragments from vehicle windscreens. *Forensic Science International* 93: 209–218.

Allen, T.J. and Scranage, J.K. (1998). The transfer of glass – Part 1: the transfer of glass to individuals at different distances. *Forensic Science International* 93: 167–174.

Amrhein, V., Greenland, S., and McShane, B. (2019). Comment: retire statistical significance. *Nature* 567: 305–307.

Anderson, T., Schum, D.A., and Twining, W. (2005). *Analysis of Evidence*, 2e. Cambridge: Cambridge University Press.

Anderson, T. and Twining, W. (1998). *Analysis of Evidence: How to Do Things with Facts Based on Wigmore's Science of Judicial Proof*. Evanston, IL: Northwestern University Press.

Antelman, G.R. (1997). *Elementary Bayesian statistics*. Cheltenham: Edward Elgar.

Arndt, J., Bell, S., Crookshanks, L., Lovejoy, M., Oleska, C., Tulley, T., and Wolfe, D. (2012). Preliminary evaluation of the persistence of organic gunshot residue. *Forensic Science International* 222: 137–145.

Arscott, E., Morgan, R., Meakin, G., and French, J. (2017). Understanding forensic expert evaluative evidence: a study of the perception of verbal expressions of strength of evidence. *Science & Justice* 57: 221–227.

Ashcroft, C.M., Evans, S., and Tebett, I.R. (1988). The persistence of fibres in head hair. *Journal of the Forensic Science Society* 28: 289–293.

Association of Forensic Science Providers (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice* 49: 161–164.

Association of Forensic Science Providers (2011). Expressing evaluative opinions: a position statement. *Science & Justice* 51: 1–2.

Ayer, M., Brunk, H.D., Ewing, G.M., Redi, W.T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* 26: 641–647.

Ayres, K.L. (2000). Relatedness testing in subdivided populations. *Forensic Science International* 114: 107–115.

Ayres, K.L. (2002). Paternal exclusion in the presence of substructure. *Forensic Science International* 129: 142–144.

Ayres, K.L. and Balding, D.J. (2005). Paternity index calculations when some individuals share common ancestry. *Forensic Science International* 151: 101–103.

Ayres, K.L. and Overall, A.D.J. (1999). Allowing for within-subpopulation inbreeding in forensic match probabilities. *Forensic Science International* 103: 207–216.

Balding, D.J. (1995). Estimating products in forensic identification using DNA profiles. *Journal of the American Statistical Association* 90: 839–844.

Balding, D.J. (1997). Errors and misunderstandings in the second NRC report. *Jurimetrics Journal* 37: 469–476.

Balding, D.J. (1999). When can a DNA profile be regarded as unique. *Science & Justice* 58: 241–244.

Balding, D.J. (2000). Interpreting DNA evidence: can probability theory help? In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 51–70. New York: Springer-Verlag.

Balding, D.J. (2002). The DNA database search controversy. *Biometrics* 58: 241–244.

Balding, D.J. (2005). *Weight-of-Evidence for Forensic DNA Profiles*. Chichester: Wiley.

Balding, D.J. and Donnelly, P. (1994). The prosecutor's fallacy and DNA evidence. *Criminal Law Review* 711–721.

Balding, D.J. and Donnelly, P. (1995a). Inference in forensic identification. *Journal of the Royal Statistical Society, Series A* 158: 21–53.

Balding, D.J. and Donnelly, P. (1995b). Inferring identity from DNA profile evidence. *Proceedings of the National Academy of Sciences of the United States of America* 92: 11741–11745.

Balding, D.J. and Donnelly, P. (1996). Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences* 41: 603–607.

Balding, D.J., Greenhalgh, M., and Nichols, R.A. (1996). Population genetics of STR loci in Caucasians. *International Journal of Legal Medicine* 108: 300–305.

Balding, D.J. and Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 64: 125–140.

Balding, D.J. and Nichols, R.A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. In: *Human Identification: The Use of DNA Markers* (ed. B.S. Weir), 3–12. Dordrecht: Kluwer Academic.

Balding, D.J. and Nichols, R.A. (1997). Significant genetic correlations among Caucasians at forensic DNA loci. *Heredity* 78: 583–589.

Balding, D.J. and Steele, C.D. (2015). *Weight-of-Evidence for Forensic DNA Profiles*, 2e. Chichester: Wiley.

Balthazard, V. (1911). De l'identification par les empreintes digitales. *Comptes rendus des seances de l'Académie des Sciences* 152: 1862–1864.

Bar-Hillel, M. and Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition* 11: 109–122.

Barnard, G.A. (1958). Thomas Bayes – a biographical note (together with a reprinting of Bayes 1763). *Biometrika* 45: 293–315. Reprinted in Pearson and Kendall (1970), 131–153.

Barnett, P.D. and Ogle, R.R. (1982). Probabilities and human hair comparison. *Journal of Forensic Sciences* 27: 272–278.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical transactions of the Royal Society of London for 1763 31: 449–455. Reprinted with Barnard (1958) in Pearson and Kendall (1970), 131–153.

Benazzi, S., Maestri, C., Parisini, S., Vecchi, F., and Gruppioni, G. (2009). Sex assessment from the sacral base by means of image processing. *Journal of Forensic Sciences* 54: 249–254.

Bentham, J. (1827). *Rationale of Judicial Evidence, Specially Applied to English Practice*. London: Hunt and Clarke.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2e. New York: Springer-Verlag.

Berger, C.E.H., Buckleton, J.S., Champod, C., Evett, I.W., and Jackson, G. (2011). Evidence evaluation: a response to the court of appeal judgment in *R v T*. *Science & Justice* 51: 43–49.

Berger, J.O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of p-values and evidence. *Journal of the American Statistical Association* 82: 112–139.

Berger, C.E.H. and Sjerps, M. (2012). Discussion paper: reaction to Hamer D. and Thompson W.C. (2012) in *Law, Probability & Risk*. *Law, Probability and Risk* 11: 373–375.

Berger, C.E.H. and Slooten, K. (2016). The LR does not exist. *Science & Justice* 56: 388–391.

Berger, C.E.H. and Stoel, R.D. (2018). Response to 'A study of the perception of verbal expressions of the strength of evidence'. *Science & Justice* 58: 76–77.

Berger, C.E.H., Vergeer, P., and Buckleton, J.S. (2015). A more straightforward derivation of the likelihood ratio for a database search. *Forensic Science International: Genetics* 14: 156–160.

Bernardo, J.M. and Smith, A.F.M. (2000). *Bayesian Theory*, 2e. Chichester: Wiley.

Bernoulli, J. (1713). *Ars conjectandi*, Basileae impensis Thurnisiorum fratrum, Basel.

Bernoulli, N. (1709). *Specimina artis conjectandi Ad quaestiones juris applicatae*, Basileae impensis Thurnisiorum fratrum, Basel.

Berry, D.A. (1991). Probability of paternity. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 150–156. Chichester: Ellis Horwood.

Berry, D.A. (1996). *A Bayesian Perspective*. Belmont, CA: Duxbury Press.

Berry, D.A., Evett, I.W., and Pinchin, R. (1992). Statistical inference in criminal investigations using deoxyribonucleic acid profiling (with discussion). *Applied Statistics* 41: 499–531.

Berry, D.A. and Geisser, S. (1986). Inference in cases of disputed paternity. In: *Statistics and the Law* (ed. M.H. DeGroot, S.E. Fienberg, and J.B. Kadane), 353–382. New York: Wiley.

Bertillon, A. (1893). *Instructions Signalétiques*. Melun: Imprimerie Administrative.

Bertillon, A. (1898). La comparaison des ecritures et l'identification graphique. In: *Revue Scientifique (December 18, 1897 – January 1, 1898)*. Paris: Typographie Chamerot et Renouard.

Bertsch McGrayne, S. (2011). *The Theory That Would Not Die*. New Haven, CT: Yale University Press.

Besson, L. (2003). Contamination en stupéfiants des billets de banque en euro. *Tech. Rep.* Université de Lausanne.

Biedermann, A., Bozza, S., Garbolino, P., and Taroni, F. (2012a). Decision-theoretic analysis of forensic sampling criteria using Bayesian decision networks. *Forensic Science International* 223: 217–227.

Biedermann, A., Taroni, F., and Champod, C. (2012b). How to assign a likelihood ratio in a footwear mark case: an analysis and discussion in the light of *R v T*. *Law, Probability and Risk* 11: 259–277.

Biedermann, A., Taroni, F., and Champod, C. (2012c). Reply to Hamer D. (2012): the *R v T* controversy: forensic evidence, law and logic. *Law, Probability and Risk* 11: 361–362.

Biedermann, A., Voisard, R., and Taroni, F. (2012d). Learning about Bayesian networks for forensic interpretation: an example based on the 'problem of multiple propositions'. *Science & Justice* 52: 191–198.

Biedermann, A., Vuille, J., and Taroni, F. (2012e). A Bayesian network approach to the database search problem in criminal proceedings. *Investigative Genetics* 3: 1–16.

Biedermann, A., Bozza, S., and Taroni, F. (2008a). Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Science International* 177: 120–132.

Biedermann, A., Taroni, F., Bozza, S., and Aitken, C.G.G. (2008b). Analysis of sampling issues using Bayesian networks. *Law, Probability and Risk* 7: 35–60.

Biedermann, A., Bozza, S., and Taroni, F. (2009a). Probabilistic evidential assessment of gunshot residue particle evidence (Part I): likelihood ratio calculation and case pre-assessment using Bayesian networks. *Forensic Science International* 191: 24–35.

Biedermann, A., Taroni, F., and Bozza, S. (2009b). Implementing statistical learning methods through Bayesian networks. (Part I): a guide to Bayesian parameter estimation using forensic science data. *Forensic Science International* 193: 63–71.

Biedermann, A., Bozza, S., and Taroni, F. (2011a). Probabilistic evidential assessment of gunshot residue particle evidence (Part II): Bayesian parameter estimation for experimental count data. *Forensic Science International* 206: 103–110.

Biedermann, A., Bozza, S., Taroni, F., and Mazzella, W.D. (2011b). Implementing statistical learning methods through Bayesian networks (Part II): Bayesian evaluations for results of black toner analyses in forensic document examination. *Forensic Science International* 204: 58–66.

Biedermann, A., Gittelson, S., and Taroni, F. (2011c). Recent misconceptions about the 'database search problem': a probabilistic analysis using Bayesian networks. *Forensic Science International* 212: 51–60.

Biedermann, A., Taroni, F., and Thompson, W.C. (2011d). Using graphical probability analysis (Bayes Nets) to evaluate conditional DNA inclusion. *Law, Probability and Risk* 10: 89–121.

Biedermann, A., Bozza, S., and Taroni, F. (2016a). The decisionalization of individualization. *Forensic Science International* 266: 29–38.

Biedermann, A., Bozza, S., Taroni, F., and Aitken, C.G.G. (2016b). Reframing the debate: a question of probability, not of likelihood ratio. *Science & Justice* 56: 392–396.

Biedermann, A., Bozza, S., Taroni, F., Fürbach, M., Li, B., and Mazzella, W.D. (2016c). Analysis and evaluation of magnetism of black toners on documents printed by electrophotographic systems. *Forensic Science International* 267: 157–165.

Biedermann, A., Champod, C., Jackson, G., Gill, P., Taylor, D., Butler, J., Morling, N., Hicks, T.N., Vuille, J., and

Taroni, F. (2016d). Evaluation of forensic DNA traces when propositions of interest relate to activities: analysis and discussion of recurrent concerns. *Frontiers in Genetics, Statistical Genetics and Methodology* 7: 1–12.

Biedermann, A., Bozza, S., Taroni, F., and Aitken, C.G.G. (2017). The consequences of understanding expert probability reporting as a decision. *Science & Justice* 57: 80–85.

Biedermann, A., Garbolino, P., and Taroni, F. (2013). The subjectivist interpretation of probability and the problem of individualization in forensic science. *Science & Justice* 53: 192–200.

Biedermann, A., Hicks, T.N., Taroni, F., Champod, C., and Aitken, C.G.G. (2014). On the use of the likelihood ratio for forensic evaluation: response to Fenton et al. Letter to the Editor. *Science & Justice* 54: 316–318.

Biedermann, A. and Taroni, F. (2006). A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues. *Forensic Science International* 163: 18–33.

Biedermann, A. and Taroni, F. (2011). Evidential relevance in scene to offender transfer cases: development and analysis of a likelihood ratio for offence level propositions. *Law, Probability and Risk* 10: 277–301.

Biedermann, A. and Taroni, F. (2012). Bayesian networks for evaluating forensic DNA profiling evidence: a review and guide to literature. *Forensic Science International: Genetics* 6: 147–157.

Biedermann, A., Taroni, F., Bozza, S., Augsburger, M., and Aitken, C.G.G. (2018). Critical analysis of forensic cut-offs and legal thresholds: a coherent approach to inference and decision. *Forensic Science International* 288: 72–80.

Biedermann, A., Taroni, F., and Garbolino, P. (2007). Equal prior probabilities: can one do any better? *Forensic Science International* 172: 85–93.

Biedermann, A. and Vuille, J. (2016). Digital evidence, 'absence' of data and ambiguous patterns of

reasonings. *Digital Investigation (DFRWS 2016 Europe (Digital Forensic Research Workshop), Proceedings of the 3rd Annual DFRWS Europe Conference*, Volume 16, Lausanne, pp. S86–S96.

Biedermann, A. and Vuille, J. (2018). The decisional nature of probability and plausibility assessments in juridical evidence and proof. *International Commentary on Evidence* 16: 1–30.

Bleka, Ø., Prieto, L., and Gill, P. (2019). CaseSolver: an investigative open source expert system based on EuroForMix. Forensic Science International: Genetics 41: 83–92.

Bodziak, W.J. (2012a). A final comment. *Law, Probability and Risk* 11: 363–364.

Bodziak, W.J. (2012b). Traditional conclusions in footwear examinations versus the use of the Bayesian approach and likelihood ratio: a review of a recent UK appellate court decision. *Law, Probability and Risk* 11: 279–287.

Bolck, A., Ni, H., and Lopatka, M. (2015). Evaluating score- and feature-based likelihood ratio models for multivariate continuous data. *Law, Probability and Risk* 14: 243–266.

Bolstad, W.M. and Curran, J.M. (2017). *Introduction to Bayesian Statistics*, 3e. Hoboken, NJ: Wiley.

Booth, G., Johnston, F., and Jackson, G. (2002). Case assessment and interpretation – application to a drugs supply case. *Science & Justice* 42: 123–125.

Bozza, S., Biedermann, A., and Taroni, F. (2018). Evaluation and reporting of scientific evidence: the impact of partial probability assignments. In: *Book of Short Papers SIS 2018* (ed. A. Abruzzo, E. Brentari, M., Chiodi, and D. Piacentino), 155–160. Pearson. https://meetings3.sis-statistica.org/index.php/sis2018/49th.

Bozza, S., Broséus, J., Esseiva, P., and Taroni, F. (2014). Bayesian classification criterion for forensic multivariate data. *Forensic Science International* 244: 295–301.

Bozza, S., Taroni, F., Marquis, R., and Schmittbuhl, M. (2008). Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. *Applied Statistics* 57: 329–341.

Breathnach, M. and Moore, E. (2015). Background levels of salivary-α-amylase plus foreign DNA in cases of oral intercourse: a female perspective. *Journal of Forensic Sciences* 60: 1563–1570.

Breathnach, M., Williams, L., McKenna, L., and Moore, E. (2016). Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. *Forensic Science International: Genetics* 20: 53–60.

Brenner, C.H. and Weir, B.S. (2003). Issues and strategies in the DNA identification of World Trade Center victims. *Theoretical Population Biology* 63: 173–178.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1–3.

Briggs, T.J. (1978). The probative value of bloodstains on clothing. *Medicine, Science and the Law* 18: 79–83.

Bright, J.A., Curran, J.M., and Buckleton, J.S. (2013). Relatedness calculations for linked loci incorporating subpopulation effects. *Forensic Science International: Genetics* 7: 380–383.

Bright, J.A., Evett, I.W., Taylor, D., Curran, J.M., and Buckleton, J.S. (2015). A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics* 14: 125–131.

Bright, J.A., Taylor, D., Kerr, Z., Buckleton, J.S., and Kruijver, M. (2019). The efficacy of DNA mixture to mixture matching. *Forensic Science International: Genetics* 41: 64–71.

Bright, J.A., Taylor, D., McGovern, C., Cooper, S., Russell, L., Abarno, D., and Buckleton, J.S. (2016). Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23: 226–239.

Bring, J. and Aitken, C.G.G. (1997). Burden of proof and estimation of drug quantities under the federal

sentencing guidelines. *Cardozo Law Review* 18: 1987–1999.

Brookfield, J.F.Y. (1994). The effect of relatives on the likelihood ratio associated with DNA profile evidence in criminal cases. *Journal of the Forensic Science Society* 34: 193–197.

Brown, G.A. and Cropp, P.L. (1987). Standardised nomenclature in forensic science. *Journal of the Forensic Science Society* 27: 393–399.

Brümmer, N. (2010). Measuring, refining and calibrating speaker and language information extracted from speech. PhD thesis. South Africa: School of Electrical Engineering, University of Stellenbosch.

Brümmer, N. and Du Preez, J. (2006). Application-independent evaluation of speaker detection. *Computer Speech & Language* 20: 230–275.

Buckleton, J.S. (1999). What can the 90's teach us about good forensic science. *1st International Conference on Forensic Human Identification in the Millenium*, London.

Buckleton, J.S. (2004). Population genetic models. In: *Forensic DNA Evidence Interpretation* (ed. C.M. Triggs, J.S. Buckleton, and S.J. Walsh), 65–118. Boca Raton, FL: CRC Press.

Buckleton, J.S., Bright, J.A., Cheng, K., Kelly, H., and Taylor, D.A. (2019). The effect of varying the number of contributors in the prosecution and alternate propositions. *Forensic Science International: Genetics* 38: 225–231.

Buckleton, J.S., Bright, J.A., Taylor, D., Evett, I.W., Hicks, T.N., Jackson, G., and Curran, J.M. (2014). Helping formulate propositions in forensic DNA analysis. *Science & Justice* 54: 258–261.

Buckleton, J.S., Bright, J.A., and Taylor, D. (2016a). Disaster victim identification, identification of missing persons and immigration cases. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 397–428. Boca Raton, FL: CRC Press.

Buckleton, J.S., Bright, J.A., and Taylor, D. (2016b). Relatedness. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 119–132. Boca Raton, FL: CRC Press.

Buckleton, J.S., Curran, J.M., Goudet, J., Taylor, D., Thiery, A., and Weir, B.S. (2016c). Population-specific F_{ST} values for forensic STR markers: a worldwide survey. *Forensic Science International: Genetics* 23: 91–100.

Buckleton, J.S., Taylor, D., and Bright, J.A. (2016d). Parentage testing. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 353–395. Boca Raton, FL: CRC Press.

Buckleton, J.S., Taylor, D., Curran, J.M., and Bright, J.A. (2016e). Population genetic models. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 87–117. Boca Raton, FL: CRC Press.

Buckleton, J.S., Taylor, D., Gill, P., Curran, J.M., and Bright, J.A. (2016f). Complex profiles. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 229–276. Boca Raton, FL: CRC Press.

Buckleton, J.S., Taylor, D., Gill, P., Curran, J.M., and Bright, J.A. (2016g). The continuous model. In: *Forensic DNA Evidence Interpretation*, 2e (ed. Buckleton, J.S., Bright, J.A. and Taylor, D.), 277–314. Boca Raton, FL: CRC Press.

Buckleton, J.S., Curran, J.M., and Walsh, S.J. (2006a). How reliable is the sub-population model in DNA testimony? *Forensic Science International* 157: 144–148.

Buckleton, J.S., Triggs, C.M., and Champod, C. (2006b). An extended likelihood ratio framework for interpreting evidence. *Science & Justice* 46: 69–78.

Buckleton, J.S. and Evett, I.W. (1989). Aspects of the Bayesian Interpretation of Fibre Evidence. *CRSE Report 684*. Home Office Forensic Science Service.

Buckleton, J.S., Evett, I.W., and Weir, B.S. (1998). Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Science & Justice* 38: 23–26.

Buckleton, J.S., Krawczak, M., and Weir, B.S. (2011). The interpretation of lineage markers in forensic DNA testing. *Forensic Science International: Genetics* 5: 78–83.

Buckleton, J.S. and Myers, S. (2014). Combining autosomal and Y chromosome match probabilities using coalescent theory. *Forensic Science International: Genetics* 11: 52–55.

Buckleton, J.S. and Triggs, C.M. (2005). Relatedness and DNA: are we taking it seriously enough? *Forensic Science International* 152: 115–119.

Buckleton, J.S., Triggs, C.M., Taroni, F., Champod, C., and Wevers, G. (2008). Experimental design for acquiring relevant data to address the issue of comparing consecutively manufactured tools and firearms. *Science & Justice* 48: 178–181.

Buckleton, J.S., Triggs, C.M., and Walsh, S.J. (2005). *Forensic DNA Evidence Interpretation*. Boca Raton, FL: CRC Press.

Buckleton, J.S. and Walsh, K.A.J. (1991). Knowledge-based systems. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 186–206. Chichester: Ellis Horwood.

Buckleton, J.S., Walsh, K.A.J., and Evett, I.W. (1991). Who is 'random man?' *Journal of the Forensic Science Society* 31: 463–468.

Buckleton, J.S., Walsh, K.A.J., Seber, G.A.F., and Woodfield, D.G. (1987). A stratified approach to the compilation of blood group frequency surveys. *Journal of the Forensic Science Society* 27: 103–112.

Budowle, B., Chakraborty, R., Carmody, G., and Monson, K.L. (2000). Source attribution of a DNA profile. *Forensic Science Communications* 2 (3): 1–6.

Bull, P.A., Morgan, R.M., Sagovsky, A., and Hughes, G.J.A. (2006). The transfer and persistence of trace

particulates: experimental studies using clothing fabrics. *Science & Justice* 46: 185–195.

Bunch, S.G. (2000). Consecutive matching striation criteria: a general critique. *Journal of Forensic Sciences* 45: 955–962.

Bunch, S.G. (2013). Application of likelihood ratios for firearms and toolmark analysis. *Science & Justice* 53: 223–229.

Burrill, J., Daniel, B., and Frascione, N. (2019). A review of trace 'touch DNA' deposits: variability factors and an exploration of cellular composition. *Forensic Science International: Genetics* 39: 8–18.

Buzzini, P. and Massonnet, G. (2013). The discrimination of colored acrylic, cotton, and wool textile fibers using Micro–Raman Spectroscopy. Part 1: In situ detection and characterization of dyes. *Journal of Forensic Sciences* 58: 1593–1600.

Buzzini, P. and Massonnet, G. (2015). The analysis of colored acrylic, cotton, and wool textile fibers using Micro–Raman Spectroscopy. Part 2: comparison with the traditional methods of fiber examination. *Journal of Forensic Sciences* 60: 712–720.

Cale, C.M., Earll, M.E., Latham, K.E., and Bush, G.L. (2016). Could secondary DNA transfer falsely place someone at the scene of a crime? *Journal of Forensic Sciences* 61: 196–202.

Caliebe, A. and Krawczak, M. (2018). Match probabilities for Y-chromosomal profiles: a paradigm shift. *Forensic Science International: Genetics* 37: 200–203.

Calman, K.C. (1996). Cancer: science and society and the communication of risk. *British Medical Journal* 313: 799–802.

Calman, K.C. and Royston, G.H.D. (1997). Risk language and dialects. *British Medical Journal* 315: 939–941.

Cammarota, V., Schnegg, M., and Massonnet, G. (2019). A study of background population of fibres on knife blades. *Forensic Science International* 296: 132–143.

Cantrell, S., Roux, C., Maynard, P., and Robertson, J. (2001). A textile fibre survey as an aid to the

interpretation of fibres evidence in the Sydney region. *Forensic Science International* 123: 48–53.

Carracedo, A., Bär, W., Lincoln, P.J., Mayr, W., Morling, N., Olaisen, B., Schneider, P.M., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G., and Wilson, M. (2000). DNA commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Science International* 110: 79–85.

Carracedo, A., Barros, F., Lareu, M.V., Pestoni, C., and Rodriguez-Calvo, M.S. (1996). Focusing the debate on forensic genetics. *Science & Justice* 36: 204–205.

Cavallini, D. and Corradi, F. (2006). Forensic identification of relatives of individuals included in a database of DNA profiles. *Biometrika* 93: 525–536.

Cereda, G. (2017). Current challenges in statistical DNA evidence evaluation. PhD thesis. The University of Lausanne, School of Criminal Justice and The University of Leiden, School of Mathematics, Lausanne and Leiden.

Cereda, G., Gill, R., and Taroni, F. (2018). A solution for the rare type match problem when using the DIP-STR marker system. *Forensic Science International: Genetics* 34: 88–96.

Chabli, S. (2001). Scene of crime evidence: fibres. In: *Proceedings of the 13th INTERPOL Forensic Science Symposium* (ed. R.E. Tontarski Jr.), 106–119. Largo, FL: National Forensic Science Technology Centre.

Champod, C. (1995). Edmond Locard, numerical standards and *probable* identification. *Journal of Forensic Identification* 45: 132–159.

Champod, C. (1996). Reconnaissance automatique et analyse statistique des minuties sur les empreintes digitales. PhD thesis. The University of Lausanne, School of Criminal Justice, Lausanne.

Champod, C. (1999). The inference of identity of source: theory and practice. *1st International Conference on Forensic Human Identification in the Millenium*, London.

Champod, C. (2000). Identification / individualization. In: *Encyclopedia of Forensic Science* (ed. J. Siegel, P.

Saukko, and G. Knupfer), 1077–1084. San Diego, CA: Academic Press.

Champod, C. (2009). Identification and individualization. In: *Wiley Encyclopedia of Forensic Science* (ed. A. Jamieson and A. Moenssens), 1508–1511. Chichester: Wiley.

Champod, C. (2013). DNA transfer: informed judgment or mere guesswork? *Frontiers in Genetics* 4: 1–3.

Champod, C., Biedermann, A., Vuille, J., Willis, S.M., and De Kinder, J. (2016a). ENFSI Guideline for evaluative reporting in forensic science: a primer for legal practitioners. *Criminal Law and Justice Weekly* 180: 189–193.

Champod, C., Lennard, C., Margot, P., and Stoilovic, M. (2016b). *Fingerprints and Other Ridge Skin Impressions*, 2e. Boca Raton, FL: CRC Press.

Champod, C. and Evett, I.W. (2000). Commentary on Broeders, A.P.A. (1999) Some observations on the use of probability scales in forensic identification. *Forensic Linguistics* 7: 228–241; *Forensic Linguistics* 7: 238–243.

Champod, C. and Evett, I.W. (2001). A probabilistic approach to fingerprint evidence. *Journal of Forensic Identification* 51: 101–122.

Champod, C. and Evett, I.W. (2009). Evidence interpretation: a logical approach. In: *Wiley Encyclopedia of Forensic Science* (ed. A. Jamieson and A. Moenssens), 968–976. Chichester: Wiley.

Champod, C., Evett, I.W., and Jackson, G. (2004). Establishing the most appropriate databases for addressing source level propositions. *Science & Justice* 44: 153–164.

Champod, C., Evett, I.W., Jackson, G., and Birkett, J. (2000). Comments on the scale of conclusions proposed by the ad hoc committee of the ENFSI marks working group. *Information Bulletin for Shoeprint/Toolmark Examiners* 6: 1–18.

Champod, C. and Jackson, G. (2001). Case assessment and Bayesian interpretation of fibres evidence.

Proceedings of the 8th Meeting of the European Fibres Group, Kracov, Poland, pp. 33–45.

Champod, C. and Jackson, G. (2002). Comments on the current debate on the Bayesian approach in marks examination. *Information Bulletin for Shoeprint/Toolmark Examiners* 8 (3): 22–25.

Champod, C. and Taroni, F. (1997). Bayesian framework for the evaluation of fibre transfer evidence. *Science & Justice* 37: 75–83.

Champod, C. and Taroni, F. (1999). Interpretation of evidence: the Bayesian approach. In: *Forensic Examination of Fibres* (ed. J. Robertson and M.C. Grieve), 379–398. London: Taylor and Francis.

Champod, C. and Taroni, F. (2017). Interpretation of evidence: the Bayesian approach. In: *Forensic Examination of Fibres*, 2e (ed. J. Robertson and C. Roux), 395–425. London: Taylor & Francis.

Champod, C., Taroni, F., and Margot, P. (1999). The Dreyfus case – an early debate on experts' conclusions (an early and controversial case on questioned document examination). *International of Forensic Document Examiners* 5: 446–459.

Chan, K.P.S. and Aitken, C.G.G. (1989). Estimation of the Bayes' factor in a forensic science problem. *Journal of Statistical Computation and Simulation* 33: 249–264.

Charpentier, A. (1993). *Historique de l'affaire Dreyfus*. Paris: Fasquelle.

Chen, X.H., Champod, C., Yang, X., Shi, S.P., Luo, Y.W., Wang, N., Wang, Y.C., and Lu, Q.M. (2019). Assessment of signature handwriting evidence via score-based likelihood ratio based on comparative measurement of relevant dynamic features. *Canadian Society of Forensic Science Journal* 52: 129–138.

Cochran, W.G. (1977). *Sampling Techniques*, 3e. Chichester: Wiley.

Cockerham, C.C. (1969). Variance of gene frequencies. *Evolution* 23: 72–84.

Cockerham, C.C. (1973). Analysis of gene frequencies. *Genetics* 74: 679–700.
Cohen, J.L. (1977). *The Probable and the Provable*. Oxford: Clarendon Press.
Cohen, J.L. (1988). The difficulty about conjunction in forensic proof. *The Statistician* 37: 415–416.
Cole, S.A. (2009). Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification. *Law, Probability and Risk* 8: 233–255.
Cole, S.A. (2014). Individualization is dead, long live individualization! Reforms of reporting practices for fingerprint analysis in the United States. *Law, Probability and Risk* 13: 117–150.
Colón, M., Rodríguez, G., and Dìaz, R.O. (1993). Representative sampling of 'street' drug exhibits. *Journal of Forensic Sciences* 38: 641–648.
Condorcet, N. (1785). *Essai sur l'application de l'analyses à la probabilité des décisions rendues a la pluralité des voix*. Paris: Imprimerie Royale.
Cook, R., Evett, I.W., Jackson, G., Jones, P.J., and Lambert, J.A. (1998a). A hierarchy of propositions: deciding which level to address in casework. *Science & Justice* 38: 231–239.
Cook, R., Evett, I.W., Jackson, G., Jones, P.J., and Lambert, J.A. (1998b). A model for case assessment and interpretation. *Science & Justice* 38: 151–156.
Cook, R., Evett, I.W., Jackson, G., and Rogers, M. (1993). A workshop approach to improving the understanding of the significance of fibres evidence. *Science & Justice* 33: 149–152.
Cook, R., Evett, I.W., Jackson, G., Jones, P.J., and Lambert, J.A. (1999). Case pre-assessment and review in a two-way transfer case. *Science & Justice* 39: 103–111.
Cook, R., Webb-Salter, M.T., and Marshall, L. (1997). The significance of fibres found in head hair. *Forensic Science International* 87: 155–160.
Cooper, G. (2013). The indirect transfer of glass fragments to a jacket and their subsequent persistence. *Science & Justice* 53: 166–170.

Corradi, F., Lago, G., and Stefanini, F.M. (2003). The evaluation of DNA evidence in pedigrees requiring population inference. *Journal of the Royal Statistical Society, Series A* 166: 425–440.

Corradi, F., Pinchi, V., Barsanti, I., and Garatti, S. (2013). Probabilistic classification of age by third molar development: the use of soft evidence. *Journal of Forensic Sciences* 58: 51–59.

Coulson, S.A., Buckleton, J.S., Gummer, A.B., and Triggs, C.M. (2001a). Glass on clothing and shoes of members of the general population and people suspected of breaking crimes. *Science & Justice* 41: 39–48.

Coulson, S.A., Coxon, A., and Buckleton, J.S. (2001b). How many samples from a drug seizure need to be analyzed? *Journal of Forensic Sciences* 46: 1456–1461.

Cournot, A. (1838). Dur les applications du calcul des chances à la statistique judiciaire. *Journal des Mathématiques Pures et Appliquées* 3: 257–334.

Cowell, R. (2016). Combining allele frequency uncertainty and population substructure corrections in forensic DNA calculations. *Forensic Science International: Genetics* 23: 210–216.

Cowell, R.G., Lauritzen, S.L., and Mortera, J. (2011). Probabilistic expert systems for handling artifacts in complex DNA mixtures. *Forensic Science International: Genetics* 5: 202–209.

Coyle, T., Jones, J., Shaw, C., and Friedrichs, R. (2012). Fibres used in the construction of car seats – an assessment of evidential value. *Science & Justice* 52: 259–267.

Cromwell, O. (1979). Letter to the general assembly of the Church of Scotland, 3rd August 1650. In: *Oxford Dictionary of Quotations*, 3e. Oxford University Press. 169.

Crupi, V., Chater, N., and Tentori, K. (2013). New axioms for probability and likelihood ratio measures. *The British Journal for the Philosophy of Science* 64: 189–204.

Crupi, V., Elia, F., Aprà, F., and Tentori, K. (2018). Double conjunction fallacies in physicians' probability judgment. *Medical Decision Making* 38: 756–760.

Cullison, A.D. (1969). *Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach*, 538–598. University of Toledo Law Review.

Curran, J.M. (2009). Use of knowledge-based systems in forensic science. In: *Wiley Encyclopedia of Forensic Science*, vol. 2 (ed. A. Jamieson and A. Moenssens), 2590–2593. Chichester: Wiley.

Curran, J.M. (2010). Are DNA profiles as rare as we think? Or can we trust DNA statistics? *Significance* 6: 62–66.

Curran, J.M. (2016). Admitting to uncertainty in the LR. *Science & Justice* 56: 380–382.

Curran, J.M. and Buckleton, J.S. (2011). An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations. *Forensic Science International: Genetics* 5: 512–516.

Curran, J.M., Buckleton, J.S., and Triggs, C.M. (2003). What is the magnitude of the subpopulation effect? *Forensic Science International* 135: 1–8.

Curran, J.M., Buckleton, J.S., Triggs, C.M., and Weir, B.S. (2002). Assessing uncertainty in DNA evidence caused by sampling effects. *Science & Justice* 42: 29–37.

Curran, J.M. and Hicks, T.N. (2009). Bayesian approach to glass evidence. In: *Wiley Encyclopedia of Forensic Science*, vol. 2 (ed. A. Jamieson and A. Moenssens), 1351–1360. Chichester: Wiley.

Curran, J.M., Hicks, T.N., and Buckleton, J.S. (2000). *Forensic Interpretation of Glass Evidence*. Boca Raton, FL: CRC Press.

Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S., and Walsh, K.A.J. (1997a). The interpretation of elemental composition measurements from forensic glass evidence: I. *Science & Justice* 37: 241–244.

Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S., and Walsh, K.A.J. (1997b). The interpretation of elemental composition measurements from forensic glass evidence: Ii. *Science & Justice* 37: 245–249.

Curran, J.M., Triggs, C.M., and Buckleton, J.S. (1998a). Sampling in forensic comparison problems. *Science & Justice* 38: 101–107.

Curran, J.M., Triggs, C.M., Buckleton, J.S., Walsh, K.A.J., and Hicks, T.N. (1998b). Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence. *Science & Justice* 38: 15–21.

Curran, J.M., Walsh, S.J., and Buckleton, J.S. (2007). Empirical testing of estimated DNA frequencies. *Forensic Science International: Genetics* 1: 267–272.

Curran, J.M., Walsh, S.J., and Buckleton, J.S. (2008). Empirical support for the reliability of DNA evidence interpretation in Australia and New Zealand. *Australian Journal of Forensic Sciences* 40: 99–108.

Dabbs, M.G.D. and Pearson, E.F. (1970). Heterogeneity in glass. *Journal of the Forensic Science Society* 10: 139–148.

Dabbs, M.G.D. and Pearson, E.F. (1972). Some physical properties of a large number of window glass specimens. *Journal of Forensic Sciences* 17: 70–78.

Daéid, N.N., McColl, D., and Ballan, J. (2009). The level of random background glass recovered from fleece jackets of individuals who worked in law enforcement or related professions. *Forensic Science International* 191: 19–23.

D'Agostini, G. (2003). Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics* 66: 138–31419.

D'Agostini, G. (2010). On the so called Boy or Girl Paradox. ArXiv:1001.0708.

D'Agostini, G. (2016). Probability, propensity and probabilities of propensities (and of probabilities). *MaxEnt*, Ghent, Belgium (15 July 2016).

Daly, D.J., Murphy, C., and McDermott, S.D. (2012). The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics* 6: 41–46.

Darboux, J.G., Appell, P.E., and Poincaré, H. (1908). Examen critique des divers systèmes ou études

graphologiques auxquels a donné lieu le bordereau. In: *L'affaire Dreyfus – La révision du procès de Rennes – Enquête de la chambre criminelle de la Cour de Cassation*. Paris: Ligue Française des droits de l'homme et du citoyen. 499–600 (original report 335–393).

Darroch, J. (1987). Probability and criminal trials: some comments prompted by the Splatt trial and the Royal Commission. *Professional Statistician* 6: 3–7.

Davis, L.J., Saunders, C.P., Hepler, A., and Buscaglia, J. (2012). Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios. *Forensic Science International* 216: 146–157.

Dawid, A.P. (1987). The difficulty about conjunction. *The Statistician* 36: 91–97.

Dawid, A.P. (2001). Comments on Stockmarr A. (1999). *Biometrics* 57: 976–980.

Dawid, A.P. (2002). Bayes's theorem and weighing evidence by juries. In: *Bayes's Theorem* (ed. R. Swinburne), 71–90. Oxford: Oxford University Press, Proceedings of the British Academy.

Dawid, A.P. (2004). Probability, causality and the empirical world: a Bayes- de Finetti- Popper- Borel synthesis. *Statistical Science* 19: 44–57.

Dawid, A.P. (2017). Forensic likelihood ratio: statistical problems and pitfalls. *Science & Justice* 57: 73–75.

Dawid, A.P. and Evett, I.W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences* 42: 226–231.

Dawid, A.P., Faigman, D.L., and Fienberg, S.E. (2014). Fitting science into legal contexts: assessing effects of causes or causes of effects? *Sociological Methods & Research* 43: 359–390.

Dawid, A.P. and Galavotti, M.C. (2009). De Finetti's subjectivism, objective probability, and the empirical validation of probability assessment. In: *Bruno de Finetti, Radical probabilist* (ed. M.C. Galavotti), 97–114. London: College Publications.

Dawid, A.P. and Mortera, J. (1996). Coherent analysis of forensic identification evidence. *Journal of the Royal Statistical Society, Series B* 58: 425–443.

Dawid, A.P., Mortera, J., and Pascali, V.L. (2001). Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. *Forensic Science International* 124: 55–61.

Dawid, A.P., Mortera, J., Pascali, V.L., and van Boxel, D. (2002). Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics* 29: 577–595.

Dawid, A.P., Musio, M., and Fienberg, S.E. (2016). From statistical evidence to evidence of causality. *Bayesian Analysis* 11: 725–752.

De Battista, R., Tidy, H., Thompson, T.J.U., and Robertson, P. (2014). An investigation into the persistence of textile fibres on buried carcasses. *Science & Justice* 54: 288–291.

de Finetti, B. (1930). Fondamenti logici del ragionamento probabilistico. *Bollettino della Unione Matematica Italiana* 9 258–261.

de Finetti, B. (1931a). Probabilismo. In: *Biblioteca di Filosofia* (ed. A. Aliotta), 1–57. Napoli: Editrice F. Perrella.

de Finetti, B. (1931b). Sul significato soggettivo della probabilità. *Fundamenta Mathematicae Italiana* XVII: 298–329.

de Finetti, B. (1933). Sul concetto di probabilità. *Rivista italiana di statistica, economia e finanza* V: 723–747.

de Finetti, B. (1940). La teoria delle probabilità nei suoi rapporti con l'analisi. *Relazioni della XXVIII Riunione della Società Italiana per il Progresso delle Scienze*, Volume terzo, Roma, Sezioni di classe A, pp. 27–35.

de Finetti, B. (1968). Probability: the subjective approach. In: *La Philosophie Contemporaine*, vol. 2 (ed. R. Klibansky), 45–53. Florence: La Nuova Italia.

de Finetti, B. (1975). *Theory of Probability*, vol. 1. London: Wiley.

de Finetti, B. (1976). La probabilità: guardarsi dalle contraffazioni. *Scientia* 111: 225–281; (English translation) Kyburg, H.E. and Smokler, H.E. (eds.) (1980). *Studies in Subjective Probability*, 2e, 194–224. New York: Dover Publications, Inc.

de Koeijer, J.A., Sjerps, M., Vergeer, P., and Berger, C.E.H. (2019). Combining evidence in complex cases – a practical approach to interdisciplinary casework. *Science & Justice*. https://doi.org/10.1016/j.scijus.2019.09.001.

De March, I., Sironi, E., and Taroni, F. (2016). Probabilistic evaluation of *n* traces with no putative source: a likelihood ratio based approach in an investigative framework. *Forensic Science International* 266: 527–533.

de Morgan, A. (1838). *An Essay on Probabilities and on Their Application to Life Contingencies and Insurance Offices*. London: London Printed for Longmans, Brown, Green and Longmans.

De Wael, K., Lepot, L., Lunstroot, K., and Gason, F. (2010). Evaluation of the shedding potential of textile materials. *Science & Justice* 50: 192–194.

de Zoete, J., Fenton, N., Noguchi, T., and Lagnado, D. (2019). Resolving the so-called 'probabilistic paradoxes in legal reasoning' with Bayesian networks. *Science & Justice* 59: 367–379.

de Zoete, J. and Sjerps, M. (2018). Combining multiple pieces of evidence using lower bound for the LR. *Law, Probability and Risk* 17: 163–178.

de Zoete, J., Sjerps, M., Lagnado, D., and Fenton, N. (2015). Modelling crime linkage with Bayesian networks. *Science & Justice* 55: 209–217.

de Zoete, J., Sjerps, M., and Meester, R. (2017). Evaluating evidence in linked crimes with multiple offenders. *Science & Justice* 57: 228–238.

DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

DeGroot, M.H. and Fienberg, S.E. (1983). The comparison and evaluation of forecasters. *The Statistician* 32: 12–22.

Devlin, B. (2000). The evidentiary value of a DNA database search. *Biometrics* 56: 1276.

Diaconis, P. and Freedman, R. (1981). The persistence of cognitive illusions. *Behavioural and Brain Sciences* 4: 333–334.

Dickson, D. (1994). As confusion leads to retrial in UK. *Nature* 367: 101–102.

Dodge, Y. (ed.) (2006). *The Oxford Dictionary of Statistical Terms*, 6e. Oxford: Oxford University Press.

Donnelly, P. and Friedman, R. (1999). DNA database searches and the legal consumption of scientific evidence. *Michigan Law Review* 97: 931–984.

Dror, I.E. (2018). Biases in forensic experts (Editorial). *Science* 360 (6386): 243.

Dujourdy, L., Barbati, G., Taroni, F., Guéniat, O., Esseiva, P., Anglada, F., and Margot, P. (2003). Evaluation of links in heroin seizures. *Forensic Science International* 131: 171–183.

Dujourdy, L., Csesztregi, T., Bovens, M., Franc, A., and Nagy, J. (2013). Sampling of illicit drugs for quantitative analysis. Part I: Heterogeneity study of illicit drugs in Europe. *Forensic Science International* 231: 249–256.

Edmond, G. (2013). Introduction – expert evidence in reports and courts. *Australian Journal of Forensic Sciences* 45: 248–262.

Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., Ballantyne, K., Searston, R.A., Thompson, M.B., Tangen, J.M., Kemp, R.I., and Martire, K.A. (2017). Thinking forensics: cognitive science for forensic practitioners. *Science & Justice* 57: 144–154.

Edwards, A.W.F. (1992). *Likelihood*, Expanded edition. Baltimore, MD: John Hopkins University Press.

Edwards, W. (1986). Comment. *Boston University Law Review* 66: 623–626.

Edwards, W. (1991). Influence diagrams, Bayesian imperialism, and the Collins case: an appeal to reason. *Cardozo Law Review* 13: 1025–1074.

Edwards, W., Lindman, H., and Savage, L.J. (1963). Bayesian statistical inference for psychological

research. *Psychological Review* 70: 193–242; Reprinted in Kadane, J.B. (ed.) (1984). *Robustness of Bayesian Analysis*. Amsterdam: Elsevier.

Eggleston, R. (1991). Similar facts and Bayes' theorem. *Jurimetrics Journal* 31: 275–287.

Egli, N., Champod, C., and Margot, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems – modelling with finger variability. *Forensic Science International* 167: 189–195.

Ellman, I.M. and Kaye, D.H. (1979). Probabilities and proof: can HLA and blood group testing prove paternity? *New York University Law Review* 54: 1131–1162.

ENFSI (2015). ENFSI Guideline for evaluative reporting in forensic science, Dublin. http://enfsi.eu/documents/forensic-guidelines/ (last accessed 03 December 2019).

ENFSI (2016). ENFSI Guideline on sampling of illicit drugs for qualitative analysis, 2nd edition, Dublin. http://enfsi.eu/documents/forensic-guidelines/ (last accessed 03 December 2019).

Engel, E. and Ventoulias, A. (1991). Monty Hall's probability puzzle. *Chance* 4: 6–9.

Essen-Möller, E. (1938). Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis: Theoretische Grundlagen. *Mitteilungen der Anthropologischen Gesellschaft* 68: 9–53.

Evett, I.W. (1977). The interpretation of refractive index measurements. *Forensic Science International* 9: 209–217.

Evett, I.W. (1978). The interpretation of refractive index measurements, II. *Forensic Science International* 12: 34–47.

Evett, I.W. (1983). What is the probability that this blood came from that person? A meaningful question? *Journal of the Forensic Science Society* 23: 35–39.

Evett, I.W. (1984). A quantitative theory for interpreting transfer evidence in criminal cases. *Journal of the Royal Statistical Society, Series C* 33: 25–32.

Evett, I.W. (1986). A Bayesian approach to the problem of interpreting glass evidence in forensic science casework. *Journal of the Forensic Science Society* 26: 3–18.

Evett, I.W. (1987a). Bayesian inference and forensic science: problems and perspectives. *The Statistician* 36: 99–105.

Evett, I.W. (1987b). On meaningful questions: a two-trace transfer problem. *Journal of the Forensic Science Society* 27: 375–381.

Evett, I.W. (1990). The theory of interpreting scientific transfer evidence. In: *Forensic Science Progress*, vol. 4 (ed. A. Maehly and R.L. Williams), 141–179. Berlin: Springer-Verlag.

Evett, I.W. (1991). Interpretation: a personal odyssey. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 9–22. New York: Ellis Horwood.

Evett, I.W. (1992). Evaluating DNA profiles in the case where the defence is 'It was my brother'. *Journal of the Forensic Science Society* 32: 5–14.

Evett, I.W. (1993a). Establishing the evidential value of a small quantity of material found at a crime scene. *Journal of the Forensic Science Society* 33: 83–86.

Evett, I.W. (1993b). Criminalistics: the future of expertise. *Journal of the Forensic Science Society* 33: 173–178.

Evett, I.W. (1995). Avoiding the transposed conditional. *Science & Justice* 35: 127–131.

Evett, I.W. (1998). Toward a uniform framework for reporting opinions in forensic science casework. *Science & Justice* 38: 198–202.

Evett, I.W. (2009). Evaluation and professionalism. *Science & Justice* 49: 159–160.

Evett, I.W. (2015). The logical foundations of forensic science: towards reliable knowledge. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370: 1–10.

Evett, I.W., Berger, C.E.H., Buckleton, J.S., Champod, C., and Jackson, G. (2017). Finding the way forward for forensic science in the US – a commentary on the PCAST report. *Forensic Science International* 278: 16–23.

Evett, I.W. and Buckleton, J. (1996). Statistical analysis of STR data. In: *Advances in Forensic Haemogenetics*, , vol. 6 (ed. A. Carracedo, B. Brinkmann, and W. Bär), 79–86. Berlin: Springer-Verlag.

Evett, I.W. and Buckleton, J.S. (1990). The interpretation of glass evidence. A practical approach. *Journal of the Forensic Science Society* 30: 215–223.

Evett, I.W., Cage, P.E., and Aitken, C.G.G. (1987). Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *Applied Statistics* 36: 174–180.

Evett, I.W., Foreman, L.A., Jackson, G., and Lambert, J.A. (2000a). DNA profiling: a discussion of issues relating to the reporting of very small match probabilities. *Criminal Law Review* 341–355.

Evett, I.W., Foreman, L.A., and Weir, B.S. (2000b). Letter to the Editor. *Biometrics* 56: 1274–1275.

Evett, I.W., Foreman, L.A., and Weir, B.S. (2000c). A response to Devlin B. (2000). *Biometrics* 56: 1277.

Evett, I.W., Jackson, G., Lambert, J., and McCrossan, S. (2000d). The impact of the principles of evidence interpretation and the structure and content of statements. *Science & Justice* 40: 233–239.

Evett, I.W., Jackson, G., and Lambert, J.A. (2000e). More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science & Justice* 40: 3–10.

Evett, I.W., Gill, P., Jackson, G., Whitaker, J., and Champod, C. (2002a). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences* 47: 520–530.

Evett, I.W., Jackson, G., Lambert, J.A., and McCrossan, S. (2002b). The impact of the principles of evidence

interpretation and the structure and content of statements. *Science & Justice* 40: 233–239.

Evett, I.W. and Lambert, J.A. (1982). The interpretation of refractive index measurements, III. *Forensic Science International* 20: 237–245.

Evett, I.W. and Lambert, J.A. (1984). The interpretation of refractive index measurements, IV. *Forensic Science International* 26: 149–163.

Evett, I.W. and Lambert, J.A. (1985). The interpretation of refractive index measurements, V. *Forensic Science International* 27: 97–110.

Evett, I.W., Lambert, J.A., and Buckleton, J.S. (1995). Further observations on glass evidence interpretation. *Science & Justice* 35: 283–289.

Evett, I.W., Lambert, J.A., and Buckleton, J.S. (1998). A Bayesian approach to interpreting footwear marks in forensic casework. *Science & Justice* 38: 241–247.

Evett, I.W., Pope, S., and Puch-Solis, R. (2016). Providing scientific guidance on DNA to the judiciary. *Science & Justice* 56: 278–281.

Evett, I.W., Scranage, J.K., and Pinchin, R. (1993). An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science. *American Journal of Human Genetics* 52: 498–505.

Evett, I.W. and Weir, B.S. (1991). Flawed reasoning in court. *Chance* 4: 19–21.

Evett, I.W. and Weir, B.S. (1998). *Interpreting DNA Evidence*. Sunderland: Sinauer Associates.

Expert Working Group on Human Factors in Latent Print Analysis (2012). *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.

Faber, N.M., Sjerps, M., Leijenhorst, H.A.L., and Maljaars, S.E. (1999). Determining the optimal sample size in forensic casework – with application to fibres. *Science & Justice* 39: 113–122.

Faigman, D.L., Kaye, D.H., Saks, M.J., and Sanders, J. (2000). How good is good enough? Expert evidence under Daubert and Kumho. *Case Western Reserve Law Review* 50: 645–667.

Fairley, W.B. (1973). Probabilistic evidence of identification evidence. *Journal of Legal Studies* 66: 493–513.

Fairley, W.B. and Mosteller, W. (1974). A conversation about Collins. *University of Chicago Law Review* 41: 242–253.

Fairley, W.B. and Mosteller, W. (1977). *Statistics and Public Policy*. London: Addison-Wesley.

Falardeau, M., Moran, V., and Muehlethaler, C. (2019). A random object-oriented population study of household paints measured by infrared spectroscopy. *Forensic Science International* 297: 72–80.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition* 43: 197–223.

Fenton, N., Berger, D., Lagnado, D., Neil, M., and Hsu, A. (2014). Response to 'On the use of the likelihood ratio for forensic evaluation': response to Fenton et al. Letter to the Editor. *Science & Justice* 54: 319–320.

Fenton, N., Neil, M., Yet, B., and Lagnado, D. (2019). Analyzing the Simonhaven case using Bayesian networks. *Topics in Cognitive Science* 1–23.

Fienberg, S.E. (1982). Statistical evidence of discrimination: comment. *Journal of the American Statistical Association* 77: 784–787.

Fienberg, S.E. (ed.) (1989). *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer-Verlag.

Fienberg, S.E. (2006). When did Bayesian inference become 'Bayesian'? *Bayesian Analysis* 1: 1–41.

Fienberg, S.E. and Finkelstein, M.P. (1996). Bayesian statistics and the law. In: *Bayesian Statistics* (ed. J.M. Bernardo and J.O. Berger, A.P. Dawid and A.F.M. Smith), vol. 5, 129–146. Oxford: Oxford University Press.

Fienberg, S.E. and Kadane, J.B. (1983). The presentation of Bayesian statistical analyses in legal proceedings. *The Statistician* 32: 88–98.

Fienberg, S.E. and Kaye, D.H. (1991). Legal and statistical aspects of some mysterious clusters. *Journal of the Royal Statistical Society, Series A* 154: 265–270.

Fienberg, S.E., Krislov, S.H., and Straf, M.L. (1996). Understanding and evaluating statistical evidence in litigation. *Jurimetrics Journal* 36: 1–32.

Fienberg, S.E. and Schervish, M.J. (1986). The relevance of Bayesian inference for the presentation of statistical evidence and for legal decision making. *Boston University Law Review* 66: 771–798.

Fimmers, R., Schneider, H., Baur, M.P., and Schneider, P.M. (2011). Erwiderung zum Brief von Taroni, et al., an den Herausgeber zum Beitrag von Schneider et al., Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNADatenbank-Treffern? (Reply to the letter of Taroni et al. to the Editor with reference to Schneider et al., Recommendations of the German Stain Commission regarding the statistical evaluation of matches following searches in the national DNA database) *Rechtsmedizin* 21: 57–60.

Finkelstein, M.O. and Fairley, W.B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review* 83: 489–517.

Finkelstein, M.O. and Fairley, W.B. (1971). A comment on 'Trial by mathematics'. *Harvard Law Review* 84: 1801–1809.

Finkelstein, M.O. and Levin, B. (2015). *Statistics for Lawyers*, 3e. New York: Springer-Verlag.

Finney, D.J. (1977). Probabilities based on circumstantial evidence. *Journal of the American Statistical Association* 72: 316–318.

Fisher, R.A. (1951). Standard calculations for evaluating a blood group system. *Heredity* 5: 51–102.

Fleming, P., Bacon, C., Blair, P., and Berry, P.J. (2000). Sudden Unexpected Deaths in Infancy. London: The Stationery Office.

Fong, W. and Inami, S.H. (1986). Results of a study to determine the probability of chance match occurrences between fibres known to be from different sources. *Journal of Forensic Sciences* 31: 65–72.

Fonneløp, A.E., Johannessen, H., Egeland, T., and Gill, P. (2016). Contamination during criminal investigation: detecting police contamination and secondary DNA transfer from evidence bags. *Forensic Science International: Genetics* 23: 121–129.

Fonneløp, A.E., Ramse, M., Egeland, T., and Gill, P. (2017). The implications of shedder status and background DNA on direct and secondary transfer in an attack scenario. *Forensic Science International: Genetics* 29: 48–60.

Fontanari, L., Gonzalez, M., Vallortigara, G., and Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences of the United States of America* 111: 17075–17080.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2010). *Statistical Distributions*, 4e. Hoboken, NJ: Wiley.

Foreman, L.A. and Evett, I.W. (2001). Statistical analysis to support forensic interpretation of a new ten-locus STR profiling system. *International Journal of Legal Medicine* 114: 147–155.

Foreman, L.A., Lambert, J.A., and Evett, I.W. (1998). Regional genetic variation in Caucasians. *Forensic Science International* 95: 27–37.

Foreman, L.A., Smith, A.F.M., and Evett, I.W. (1997). A Bayesian approach to validating STR multiplex databases for use in forensic casework. *International Journal of Legal Medicine* 110: 244–250.

Forensic Science Regulator (2012). Report into the circumstances of a complaint received from the Greater

Manchester Police on 7 March 2012 regarding DNA evidence provided by LGC Forensics – FSR-R-618. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/118941/dna-contam-report.pdf (accessed 08 June 2020).

Frank, R.S., Hinkley, S.W., and Hoffman, C.G. (1991). Representative sampling of drug seizures in multiple containers. *Journal of Forensic Sciences* 36: 350–357.

Franklin, J. (2016). Pre-history of probability. In: *The Oxford Handbook of Probability and Philosophy* (ed. A. Hajek and C. Hitchcock), 33–49. Oxford: Oxford University Press.

French, J.C. and Morgan, R.M. (2015). An experimental investigation of the indirect transfer and deposition of gunshot residue: further studies carried out with SEM-EDX analysis. *Forensic Science International* 247: 14–17.

French, J.C., Morgan, R.M., Baxendell, P., and Bull, P.A. (2012). Multiple transfers of particulates and their dissemination within contact networks. *Science & Justice* 52: 33–41.

Friedman, R.D. (1986a). A close look at probative value. *Boston University Law Review* 66: 733–759.

Friedman, R.D. (1986b). A diagrammatic approach to evidence. *Boston University Law Review* 66: 571–622.

Friedman, R.D. (1996). Assessing evidence. *Michigan Law Review* 94: 1810–1838.

Friedman, R.D., Kaye, D.H., Mnookin, J., Nance, D., and Saks, M.J. (2002). Expert testimony on fingerprints: an internet exchange. *Jurimetrics Journal* 43: 91–98.

Fung, W.K. (2003). User-friendly programs for easy calculations in paternity testing and kinship determinations. *Forensic Science International* 136: 22–34.

Fung, W.K., Carracedo, A., and Hu, Y.Q. (2003). Testing for kinship in a subdivided population. *Forensic Science International* 135: 105–109.

Fung, W.K. and Hu, Y.Q. (2008). *Statistical DNA Forensics: Theory, Methods and Computation*. Chichester: Wiley.

Galavotti, M.C. (2016). The origin of probabilistic epistemology – some leading 20th-century philosophers of probability. In: *The Oxford Handbook of Probability and Philosophy* (ed. A. Hajek and C. Hitchcock), 130–151. Oxford: Oxford University Press.

Galavotti, M.C. (2017). The interpretation of probability: still an open issue? *Philosophies* 2: 1–13.

Gallidabino, M., Biedermann, A., and Taroni, F. (2015). Commentary on: Gauriot et al. (2013). *Journal of Forensic Sciences* 60: 539–541.

Garber, D. and Zabell, S. (1979). On the emergence of probability. *Archive for History of Exact Sciences* 21: 33–53.

Garbolino, P. (2001). Explaining relevance. *Cardozo Law Review* 22: 1503–1521.

Garbolino, P. (2014). *Probabilità e logica della prova*. Milano: Giuffré Editore.

Garbolino, P. and Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks. *Forensic Science International* 125: 149–155.

Gastwirth, J.L. (1998a). *Statistical Reasoning in Law and Public Policy: Statistical Concepts and Issues of Fairness*, vol. 1. San Diego, CA: Academic Press.

Gastwirth, J.L. (1998b). *Statistical Reasoning in Law and Public Policy: Tort Law, Evidence and Health*, vol. 2. San Diego, CA: Academic Press.

Gastwirth, J.L. (ed.) (2000). *Statistical Science in the Courtroom*. New York: Springer-Verlag.

Gastwirth, J.L. (2017). Some recurrent problems in interpreting statistical evidence in equal employment cases. *Law, Probability and Risk* 16: 181–201.

Gastwirth, J.L., Freidlin, B., and Miao, W. (2000). The *Shonubi* case as an example of the legal system's failure to appreciate statistical evidence. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 405–413. New York: Springer-Verlag.

Gaudette, B.D. (1982). A supplementary discussion of probabilities and human hair comparisons. *Journal of Forensic Sciences* 27: 279–289.

Gaudette, B.D. (1986). Evaluation of associative physical evidence. *Journal of the Forensic Science Society* 26: 163–167.

Gaudette, B.D. (1999). Evidential value of hair examination. In: *Forensic Examination of Hair* (ed. J. Robertson), 243–260. London: Taylor & Francis.

Gaudette, B. (2000). Comparison: significance of hair comparison. In: *Encyclopedia of Forensic Science* (ed. J.A. Siegel, P.J. Saukko, and G.C. Knupfer), 1018–1024. San Diego, CA: Academic Press.

Gaudette, B.D. and Keeping, E.S. (1974). An attempt at determining probabilities in human scalp hair comparison. *Journal of Forensic Sciences* 19: 599–606.

Gauriot, R., Gunaratnam, L., Moroni, R., Reinikainen, T., and Corander, J. (2013). Statistical challenges in the quantification of gunshot residue evidence. *Journal of Forensic Sciences* 58: 1149–1155.

Geisser, S. (1993). *Predictive Inference: An Introduction*. London: Chapman & Hall.

Gelfand, A.E. and Solomon, H. (1973). A study of Poisson's models for jury verdicts in criminal and civil trials. *Journal of the American Statistical Association* 68: 271–278.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2014). *Bayesian Data Analysis*, 3e. Boca Raton, FL: CRC Press.

Gettinby, G. (1984). An empirical approach to estimating the probability of innocently acquiring bloodstains of different ABO groups in clothing. *Journal of the Forensic Science Society* 24: 221–227.

Gill, P. (2016). Analysis and implications of the miscarriages of justice of Amanda Knox and Raffaele Sollecito. *Forensic Science International: Genetics* 23: 9–18.

Gill, P. (2019). DNA evidence and miscarriages of justice. *Forensic Science International* 294: e1–e3.

Gill, P., Hicks, T.N., Butler, J., Connolly, E., Gusmão, L., Kokshoorn, B., Morling, N., van Oorschot, R.A.H., Parson, W., Prinz, M., Schneider, P.M., Sijen, T., and Taylor, D. (2018). DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biological evidence – Guidelines highlighting the importance of propositions. Part I: Evaluation of DNA profiling comparisons given (sub-)source propositions. *Forensic Science International: Genetics* 36: 189–202.

Gill, P., Hicks, T.N., Butler, J., Connolly, E., Gusmão, L., Kokshoorn, B., Morling, N., van Oorschot, R.A.H., Parson, W., Prinz, M., Schneider, P.M., Sijen, T., and Taylor, D. (2020). DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biological evidence – Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions. *Forensic Science International: Genetics* 44. https://doi.org/10.1016/j.fsigen.2019.102186.

Girotto, V. and Pighin, S. (2015). Basic understanding of posterior probability. *Frontiers in Psychology* 6: 1–3.

Gittelson, S. (2013). Evolving from inferences to decisions in forensic science. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Gittelson, S., Berger, C.E.H., Jackson, G., Evett, I.W., Champod, C., Robertson, B., Curran, J.M., Taylor, D., Weir, B.S., Coble, M.D., and Buckleton, J.S. (2018). A response to 'Likelihood ratio as weight of evidence: a closer look' by Lund and Iyer. *Forensic Science International* 288: e15–e19.

Gittelson, S., Biedermann, A., Bozza, S., and Taroni, F. (2012a). Bayesian networks and the value of the evidence for the forensic two-trace transfer problem. *Journal of Forensic Sciences* 57: 1199–1216.

Gittelson, S., Biedermann, A., Bozza, S., and Taroni, F. (2012b). The database search problem: a question of rational decision making. *Forensic Science International* 222: 186–199.

Gittelson, S., Biedermann, A., Bozza, S., and Taroni, F. (2013a). Modeling the forensic two-trace problem with Bayesian networks. *Artificial Intelligence and the Law* 21: 221–252.

Gittelson, S., Bozza, S., Biedermann, A., and Taroni, F. (2013b). Decision-theoretic reflections on processing a fingermark. *Forensic Science International* 226: e42–e47.

Gittelson, S., Biedermann, A., Bozza, S., and Taroni, F. (2014). Decision analysis for the genotype designation in low-template-DNA profiles. *Forensic Science International: Genetics* 9: 118–133.

Gittelson, S., Kalafut, T., Myers, S., Taylor, D., Hicks, T.N., Taroni, F., Evett, I.W., Bright, J.A., and Buckleton, J.S. (2016a). A practical guide for the formulation of propositions in the Bayesian approach to the DNA evidence interpretation in an adversarial environment. *Journal of Forensic Sciences* 61: 186–195.

Gittelson, S., Steffen, C.R., and Coble, M.D. (2016b). Low-template DNA: a single DNA analysis or two replicates? *Forensic Science International* 264: 139–145.

Gjertson, D.W., Brenner, C.H., Baur, M.P., Carracedo, A., Guidet, F., Luque, J.A., Lessig, R., Mayrh, W.R., Pascali, V.L., Prinz, M., Schneider, P.M., and Morling, N. (2007). ISFG: recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics* 1: 223–231.

Goldmann, T., Taroni, F., and Margot, P. (2004). Analysis of dyes in illicit pills (amphetamine and derivatives). *Journal of Forensic Sciences* 49 (4): 1–7.

Gonzalez-Rodriguez, J., Drygajlo, A., Ramos-Castro, D., Garcia-Gomar, M., and Ortega-Garcia, J. (2006). Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. *Computer Speech & Language* 20 (2–3): 331–355.

Gonzalez-Rodriguez, J., Fierrez-Aguilar, J., Ramos-Castro, D., and Ortega-Garcia, J. (2005). Bayesian analysis of

fingerprint, face and signature evidences with automatic biometric systems. *Forensic Science International* 155: 126–140.

Gonzalez-Rodriguez, J., Rose, P., Ramos-Castro, D., Toledano, D.T., and Ortega-Garcia, J. (2007). Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 15: 2104–2115.

Good, I.J. (1950). *Probability and the Weighing of Evidence*. London: Griffin.

Good, I.J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B* 14: 107–114.

Good, I.J. (1956). Discussion of paper by G. Spencer Brown. In: *Information Theory, Third London Symposium 1955* (ed. C. Cherry), 13–14. London: Butterworths.

Good, I.J. (1959). Kinds of probability. *Science* 129: 443–447.

Good, I.J. (1979). Studies in the history of probability and statistics XXXVII: A.M. Turing's statistical work in World War II. *Biometrika* 66: 393–396.

Good, I.J. (1983). A correction concerning my interpretation of Peirce and the Bayesian interpretation of Neyman-Pearson hypothesis determination. *Journal of Statistical Computation and Simulation* 18: 71–74.

Good, I.J. (1985). Weight of evidence: a brief survey. In: *Bayesian Statistics*, vol. 2 (ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), 249–270. Amsterdam: North-Holland.

Good, I.J. (1989). C319: weight of evidence and a compelling metaprinciple. *Journal of Statistical Computation and Simulation* 31: 121–123.

Good, I.J. (1991). Weight of evidence and the Bayesian likelihood ratio. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 85–106. Chichester: Ellis Horwood.

Goodman, J. (1992). Jurors' comprehension and assessment of probabilistic evidence. *American Journal of Trial Advocacy* 16: 361.

Graversen, T., Mortera, J., and Lago, G. (2019). The Yara Gambirasio case: combining evidence in a complex DNA mixture. *Forensic Science International: Genetics* 40: 52–63.

Green, P.J. and Mortera, J. (2017). Paternity testing and other inference about relationships from DNA mixtures. *Forensic Science International: Genetics* 28: 128–137.

Grieve, M.C. (2001). A survey on the evidential value of fibres and on the interpretation of the findings in fibre transfer cases. Part 2: Interpretation and reporting. *Science & Justice* 40: 201–209.

Grieve, M.C. and Dunlop, J. (1992). A practical aspect of the Bayesian interpretation of fibre evidence. *Journal of the Forensic Science Society* 32: 169–175.

Grieve, M.C., Roux, C., and Wiggins, K.G. (2017). Factors influencing interpretation. In: *Forensic Examination of Fibres* (ed. J. Robertson and C. Roux), 346–375. London: Taylor & Francis.

Grima, M., Butler, M., Hanson, R., and Mohameden, A. (2012). Firework displays as sources of particles similar to gunshot residue. *Science & Justice* 52: 49–57.

Groom, P.S. and Lawton, M.E. (1987). Are they a pair? *Journal of the Forensic Science Society* 27: 189–192.

Grove, D.M. (1980). The interpretation of forensic evidence using a likelihood ratio. *Biometrika* 67: 243–246.

Grove, D.M. (1981). The statistical interpretation of refractive index measurements. *Forensic Science International* 18: 189–194.

Grove, D.M. (1984). The statistical interpretation of refractive index measurements II: the multiple source problem. *Forensic Science International* 24: 173–182.

Gunel, E. and Wearden, S. (1995). Bayesian estimation and testing of gene frequencies. *Theoretical and Applied Genetics* 91: 534–543.

Gusmão, L., Butler, J., Carracedo, A., Gill, P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., and Schneider, P.M. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *International Journal of Legal Medicine* 120: 191–200.

Haaf, J.M., Ly, A., and Wagenmakers, E.J. (2019). Retire significance, but still test hypotheses. *Nature* 567: 461.

Habbema, J.D.F., Hermans, J., and van den Broek, K. (1974). A stepwise discrimination program using density estimation. In: *Compstat 1974* (ed. G. Bruckman), 100–110. Vienna: Physica Verlag.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

Halliwell, J., Keppens, J., and Shen, Q. (2003). Linguistic Bayesian networks for reasoning with subjective probabilities in forensic statistics. *9th International Conference on Artificial Intelligence and Law (ICAIL 2003)*, Edinburgh, pp. 42–50.

Hamer, D. (2012). Discussion paper: the *R v T* controversy: forensic evidence, law and logic. *Law, Probability and Risk* 11: 331–345.

Hannigan, T.J., McDermott, S.D., Greaney, C.M., O'Shaughnessy, J., and O'Brien, C.M. (2015). Evaluation of gunshot residue (GSR) evidence: surveys of prevalence of GSR on clothing and frequency of residue types. *Forensic Science International* 257: 177–181.

Harbison, S.A. and Buckleton, J.S. (1998). Applications and extensions of subpopulation theory: a caseworkers' guide. *Science & Justice* 38: 249–254.

Harbison, S.A., Stanfield, A.M., Buckleton, J.S., and Walsh, S.J. (2002). Allele frequencies for four major sub-populations in New Zealand at three STR loci – HUMTH01, HUMTPOX and CSF1PO. *Forensic Science International* 27: 171–187.

Harrison, P.H., Lambert, J.A., and Zoro, J.A. (1985). A survey of glass fragments recovered from clothing of persons suspected of involvement in crime. *Forensic Science International* 27: 171–187.

Harvey, W., Butler, O., Furness, J., and Laird, R. (1968). The Biggar murder: dental, medical, police and legal aspects. *Journal of the Forensic Science Society* 8: 155–219.

Henry, E.R. (1913). *Classification and Uses of Finger Prints*, 4e. London: Darling & Sons (First edition 1900).

Hepler, A.B., Dawid, A.P., and Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk* 6: 275–293.

Hepler, A.B., Saunders, C.P., Davis, L.J., and Buscaglia, J. (2012). Score-based likelihood ratios for handwriting evidence. *Forensic Science International* 219: 129–140.

Hepler, A.B. and Weir, B.S. (2008). Object-oriented Bayesian networks for paternity cases with allelic dependencies. *Forensic Science International: Genetics* 2: 166–175.

Hicks, T.N., Biedermann, A., de Koeijer, J.A., Taroni, F., Champod, C., and Evett, I.W. (2015). The importance of distinguishing information from evidence/observations when formulating propositions. *Science & Justice* 55: 520–525.

Hicks, T.N., Buckleton, J.S., Bright, J.A., and Taylor, D. (2016). A framework for interpreting evidence. In: *Forensic DNA Evidence Interpretation*, 2e (ed. J.S. Buckleton, J.A. Bright, and D. Taylor), 37–86. Boca Raton, FL: CRC Press.

Hicks, T.N., Schütz, F., Curran, J.M., and Triggs, C.M. (2005). A model for estimating the number of glass fragments transferred when breaking a pane: experiments with firearms and hammer. *Science & Justice* 45: 65–74.

Hill, R. (2004). Multiple sudden infant deaths – coincidence or beyond coincidence? *Paediatric and Perinatal Epidemiology* 18: 320–226.

Hodges, A. (1992). *Alan Turing: Enigma*. London: Vintage Books.

Hofstetter, C., Maitre, M., Beavis, A., Roux, C., Weyermann, C., and Gassner, A.L. (2017). A study of transfer and prevalence of organic gunshot residues. *Forensic Science International* 277: 241–251.

Holden, C. (1997). DNA fingerprinting comes of age. *Science* 278: 1407.

Hong, S., Han, A., Kim, S., Son, D., and Min, H. (2014). Transfer of fibres on the hands of living subjects and their persistence during hand washing. *Science & Justice* 54: 451–458.

Hopwood, A.J., Puch-Solis, R., Tucker, V.C., Curran, J.M., Skerrett, J., Pope, S., and Tully, G. (2012). Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts. *Science & Justice* 52: 185–190.

Howard, R.A. and Matheson, J.E. (1984). Influence diagrams. In: *Readings on the Principles and Applications of Decision Analysis*, vol. 2 (ed. R.A. Howard and J.E. Matheson), 719–762. Menlo Park, CA: Strategic Decisions Group.

Howson, C. and Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach*, 2e. Chicago: Open Court Publishing Company.

Hummel, K. (1971). Biostatistical opinion of parentage based upon the results of blood group tests. In: *Biostatistische Abstammungsbegutachtung mit Blutgruppenbefunden* (ed. P. Schmidt). Stuttgart: Gustav Fisher. (Quoted in *Family Law Quarterly* 10 (1976). 262).

Hummel, K. (1983). Selection of gene frequency tables. In: *Inclusion Probabilities in Parentage Testing* (ed. R.H. Walker), 231–243. Arlington, VA: American Association of Blood Banks.

Ihaka, R. and Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.

Imwinkelried, E.J. (2001). Can we rely on the alleged constitutional right to informational privacy to secure genetic privacy in the courtroom? *Seton Hall Law Review* 31: 926–935.

Imwinkelried, E.J. (2016). Computer source code: a source of the growing controversy over the reliability of automated forensic techniques. *DePaul Law Review* 66: 97–132.

Inman, K.I. and Rudin, N. (2001). *Principles and Practice of Criminalistics – The Profession of Forensic Science*. Boca Raton, FL: CRC Press.

Ioannidis, J.P.A. (2019). Retiring statistical significance would give bias a free pass. *Nature* 567: 461.

Irwin, M. (2011). Transfer of glass fragments when bottles and drinking glasses are broken. *Science & Justice* 51: 16–18.

Izenman, A.J. (2000a). Assessing the statistical evidence in the *Shonubi* case. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 415–433. New York: Springer-Verlag.

Izenman, A.J. (2000b). Introduction to two views on the *Shonubi* case. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 393–403. New York: Springer-Verlag.

Izenman, A.J. (2000c). Statistical issues in the application of the Federal sentencing guidelines in drug, pornography and fraud cases. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 25–50. New York: Springer-Verlag.

Izenman, A.J. (2001). Statistical and legal aspects of forensic study of illicit drugs. *Statistical Science* 16: 35–57.

Izenman, A.J. (2003). Sentencing illicit drug traffickers: how do the courts handle random sampling issues? *International Statistical Review* 71: 535–556.

Jackson, G. (2000). The scientist and the scales of justice. *Science & Justice* 40: 81–85.

Jackson, G. (2013). The impact of commercialization on the evaluation of DNA evidence. *Frontiers in Genetics* 4: 1–3.

Jackson, G., Aitken, C.G.G., and Roberts, P. (2014). *Case Assessment and Interpretation of Expert Evidence* (Practitioner Guide No. 4). London: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society.

Jackson, G. and Jones, P.J. (2009). Case assessment and interpretation. In: *Wiley Encyclopedia of Forensic Science* (ed. A. Jamieson and A. Moenssens), 483–497. Chichester: Wiley.

Jackson, G., Jones, S., Booth, G., Champod, C., and Evett, I.W. (2006). The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings. *Science & Justice* 46: 33–44.

Jackson, F., Maynard, P., Cavanagh-Steer, K., Dusting, T., and Roux, C. (2013). A survey of glass found on the headwear and head hair of a random population *vs.* people working with glass. *Forensic Science International* 226: 125–131.

Jacquet, M. and Champod, C. (2020). Automated face recognition in forensic science: review and perspective. *Forensic Science International* 307: 110–124.

Jamieson, A. (2016). Letter: commentary on Breathnach et al. *Forensic Science International: Genetics* 25: e4–e5.

Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffrey, R.C. (1975). Probability and falsification: critique of the Popper program. *Synthesè* 30: 95–117.

Jeffrey, R.C. (1983). *The Logic of Decision*, 2e. Chicago: University of Chicago Press.

Jeffreys, H. (1983). *Theory of Probability*, 3e. Oxford: Clarendon Press.

Jeffreys, A.J., Wilson, V., and Morton, D.B. (1987). DNA fingerprints of dogs and cats. *Animal Genetics* 18: 1–15.

Jensen, F.V. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag.

Jevons, W.S. (1913). *The Principles of Science – A Treatise on Logic and Scientific Method (First Edition 1874)*. London: MacMillan and Co.

Johnson, V.E. (2019). Raise the bar rather than retire significance. *Nature* 567: 461.

Johnson, R.E. and Peterson, J. (1999). HLA-DQA1 and polymarker locus allele frequencies for Chicago, Illinois, U.S.A. *Journal of Forensic Sciences* 44: 1097.

Jolliffe, I.T. (1986). *Principal Component Analysis*. New York: Springer-Verlag.

Jones, D.A. (1972). Blood samples: probability of discrimination. *Journal of the Forensic Science Society* 12: 355–359.

Jones, S., Scott, K., Lewis, J., Davidson, G., and Baird, A. (2016). DNA transfer through nonintimate social contact. *Science & Justice* 56: 90–95.

Jordan, M.I. (ed.) (1999). *Learning in Graphical Models*. Cambridge: The MIT Press.

Juchli, P. (2016). Combining evidence. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Juchli, P., Biedermann, A., and Taroni, F. (2012). Graphical probabilistic analysis of the combination of items of evidence. *Law, Probability and Risk* 11: 51–84.

Kadane, J.B. (ed.) (2008). *Statistics in the Law*. Oxford: Oxford University Press.

Kadane, J.B. (2011). *Principles of Uncertainty*. Boca Raton, FL: Chapman & Hall.

Kadane, J.B. (2018a). *Batson* and reverse-*Batson* motions in North Carolina: *State v. Hurd and State v. Tucker*. *Law, Probability and Risk* 17: 263–273.

Kadane, J.B. (2018b). Statistics for *Batson* challenges. *Law, Probability and Risk* 17: 1–13.

Kadane, J.B. and Schum, D.A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. New York: Wiley.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kaplan, T. (2014). Monte Sereno murder case casts doubts on DNA evidence. San Jose Mercury News (28 June).

Kass, R.E. and Raftery, A.E. (1995). Bayes' factors. *Journal of the American Statistical Association* 90: 773–795.

Katterwe, H. (2002a). Comments/objections to reproaches of the Forensic Science Service and the University of Lausanne. *Information Bulletin for Shoeprint/Toolmark Examiners* 8 (1): 25–30.

Katterwe, H. (2002b). Comments of Horst Katterwe to the article of Taroni and Buckleton (2002). *Information Bulletin for Shoeprint/Toolmark Examiners* 8 (3): 16–20.

Katterwe, H. (2003). True or false. *Information Bulletin for Shoeprint/Toolmark Examiners* 9 (2): 18–25.

Kaye, D.H. (1979). The laws of probability and the law of the land. *University of Chicago Law Review* 47: 34–56.

Kaye, D.H. (1982). Statistical evidence of discrimination. *Journal of the American Statistical Association* 77: 773–783.

Kaye, D.H. (1986a). Is proof of statistical significance relevant? *Washington Law Review* 61: 1333–1365.

Kaye, D.H. (1986b). Quantifying probative value. *Boston University Law Review* 66: 761–766.

Kaye, D.H. (1987a). The admissibility of 'probability evidence' in criminal trials – Part II. *Jurimetrics Journal* 27: 160–172.

Kaye, D.H. (1987b). Apples and oranges: confidence coefficients and the burden of persuasion. *Cornell Law Review* 73: 54–77.

Kaye, D.H. (1987c). The validity of tests: caveant omnes. *Jurimetric Journal* 27: 349–361.

Kaye, D.H. (1989). The probability of an ultimate issue: the strange cases of paternity testing. *Iowa Law Review* 75: 75–109.

Kaye, D.H. (1993). DNA evidence: probability, population genetics, and the courts. *Harvard Journal of Law and Technology* 7: 101–172.

Kaye, D.H. (1993). Proceedings of the Second International Conference on Forensic Statistics, Tech. Rep., Arizona State University, Center for the Study of Law, Science and Technology, Tempe, Arizona, USA. *Selected papers included in Jurimetrics Journal* 34 (1): 11–15.

Kaye, D.H. (1997). DNA identification in criminal cases: some lingering and emerging evidentiary issues. *Proceedings of the 7th International Symposium on Human Identification*. Madison, Wisconsin, USA: Promega Corporation, pp. 12–25.

Kaye, D.H. (2001). Bioethical objections to DNA databases for law enforcement: questions and answers. *Seton Hall Law Review* 31: 936–948.

Kaye, D.H. (2004). Logical relevance: problems with the reference population and DNA mixtures in *People v. Pizaro*. *Law, Probability and Risk* 3: 211–220.

Kaye, D.H. (2007). Revisiting 'Dreyfus': a more complete account of a trial by mathematics. *Minnesota Law Review* 91: 825–835.

Kaye, D.H. (2008a). DNA probabilities in *People v. Prince*: when are racial and ethnic statistics relevant? In: *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2 (ed. D. Nolan and T. Speed), 289–301. Beachwood, OH: Institute of Mathematical Statistics Collections.

Kaye, D.H. (2008b). The role of race in DNA statistics: what experts say, what California courts allow. *Southwestern University Law Review* 37: 304–322.

Kaye, D.H. (2009a). Identification, individualization and uniqueness: what is the difference? *Law, Probability and Risk* 8: 85–94.

Kaye, D.H. (2009b). Rounding up the usual suspects: a logical and legal analysis of DNA trawling cases. *North Carolina Law Review* 87: 425–503.

Kaye, D.H. (2009c). Trawling, DNA databases for partial matches: what is the FBI afraid of? *Cornell Journal of Law and Public Policy* 19: 145–171.

Kaye, D.H. (2011). The expected value fallacy in *State v. Wright*. *Jurimetrics Journal* 52: 1–6.

Kaye, D.H. (2013). Beyond uniqueness: the birthday paradox, source attribution and individualization in forensic science testimony. *Law, Probability and Risk* 12: 3–11.

Kaye, D.H. (2015). Ultracrepidarianism in forensic science: the hair evidence debacle. *Washington and Lee Law Review Online* 72: 227–254.

Kaye, D.H. (2016). The interpretation of DNA evidence: a case study in probabilities – an educational module, Washington, DC. https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_173207.pdf (last accessed 03 December 2019).

Kaye, D.H. (2017a). Deadly statistics: quantifying an 'unacceptable risk' in capital punishment. *Law, Probability and Risk* 16: 7–34.

Kaye, D.H. (2017b). Digging into the foundations of evidence law. *Michigan Law Review* 115: 915–934.

Kaye, D.H. (2017c). Forensic commentary series. Hypothesis testing in law and forensic science: a memorandum. *Harvard Law Review Forum* 130 (5): 127–136.

Kaye, D.H. and Aickin, M. (eds.) (1986). *Statistical Methods in Discrimination Litigation*. New York: Marcel Dekker.

Kaye, D.H. and Freedman, D.A. (2002). Statistical proof. In: *Modern Scientific Evidence – The Law and Science of Expert Testimony*, 2e, (ed. D.L. Faigman), vol. 1, 155–246. St. Paul, MN: West Publishing Co..

Kaye, D.H. and Koehler, J.J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society, Series A* 154: 75–81.

Kaye, D.H. and Koehler, J.J. (2003). The misquantification of probative value. *Law and Human Behavior* 27: 645–659.

Kaye, D.H. and Sensabaugh, G.F. (2000). Reference guide on DNA evidence. In: *Reference Manual on Scientific Evidence* (ed. J. Cecil), 485–576. Washington, DC: Federal Judicial Center.

Kelly, H., Bright, J.A., Kruijver, M., Cooper, S., Taylor, D., Duke, K., Strong, M., Beamer, V., Buettner, C., and Buckleton, J.S. (2018). A sensitivity analysis to determine the robustness of STRmix with respect to laboratory calibration. *Forensic Science International: Genetics* 35: 113–122.

Kennedy, R.B., Pressman, I.S., Chen, S., Petersen, P.H., and Pressman, A.E. (2003). Statistical analysis of barefoot impressions. *Journal of Forensic Sciences* 48: 55–63.

Kerkvliet, T. and Meester, R. (2016). Assessing forensic evidence by computing belief functions. *Law, Probability and Risk* 15: 127–153.

Kharbouche, H., Faouzi, M., Sanchez, J.B., Daeppen, N., Augsburger, M., Mangin, P., Staub, C., and Sporkert, F. (2012). Diagnostic performance of ethyl glucoronide in hair for the investigation of alcohol drinking behaviour: a comparison with traditional biomarkers. *International Journal of Legal Medicine* 126: 243–250.

Kind, S.S. (1994). Crime investigation and the criminal trial: a three chapter paradigm. *Journal of the Forensic Science Society* 34: 155–164.

Kind, S.S., Wigmore, R., Whitehead, P.H., and Loxley, D.S. (1979). Terminology in forensic science. *Journal of the Forensic Science Society* 19: 189–192.

Kingston, C.R. (1965a). Applications of probability theory in criminalistics. *Journal of the American Statistical Association* 60: 70–80.

Kingston, C.R. (1965b). Applications of probability theory in criminalistics – II. *Journal of the American Statistical Association* 60: 1028–1034.

Kingston, C.R. (1966). Probability and legal proceedings. *The Journal of Criminal Law, Criminology and Police Science* 57: 93–98.

Kingston, C.R. (1988). Discussion of 'A critical analysis of quantitative fingerprint individuality models'. *Journal of Forensic Sciences* 33: 9–11.

Kingston, C.R. and Kirk, P.L. (1964). The use of statistics in criminalistics. *Journal of Criminal Law, Criminology and Police Science* 55: 514–521.

Kirk, P.L. (1963). The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science* 54: 235–238.

Kirk, P.L. and Kingston, C.R. (1964). Evidence evaluation and problems in general criminalistics. *Journal of Forensic Sciences* 9: 434–444.

Koehler, J.J. (1992). Probabilities in the courtroom: an evaluation of the objection and policies. In: *Handbook of Psychology and Law* (ed. D.K. Kagehiro and W.S. Laufer), 167–184. New York: Springer-Verlag.

Koehler, J.J. (1993a). Error and exaggeration in the presentation of DNA evidence: irrelevant and predjudicial. *Jurimetrics Journal* 34: 21–39.

Koehler, J.J. (1993b). DNA matches and statistics: important questions, surprising answers. *Judicature* 76: 222–229.

Koehler, J.J. (1996). On conveying the probative value of DNA evidence: frequencies, likelihood ratios, and error rates. *University of Colorado Law Journal* 67: 859–886.

Koehler, J.J. (1997a). One in millions, billions, and trillions: lessons from *People v. Collins* (1968) for *People v. Simpson* (1995). *Journal of Legal Education* 47: 214–223.

Koehler, J.J. (1997b). Why DNA likelihood ratios should account for error (even when a National Research Council report says they should not). *Jurimetrics Journal* 37: 425–437.

Koehler, J.J. (2001a). The psychology of numbers in the courtroom: how to make DNA match statistics seem impressive or insufficient. *Southern California Law Review* 74: 1275–1306.

Koehler, J.J. (2001b). When are people persuaded by DNA match statistics? *Law and Human Behaviour* 25: 493–513.

Koehler, J.J. (2008). Fingerprint error rates and proficiency tests: what they are and why they matter. *Hastings Law Journal* 59: 1077–1099.

Koehler, J.J. (2013). Proficiency tests to estimate error rates in the forensic sciences. *Law, Probability and Risk* 12: 89–98.

Koehler, J.J. (2014). Forensic fallacies and a famous judge. *Jurimetrics Journal* 54: 211–219.

Koehler, J.J. (2017a). Forensics or Fauxrensics? Ascertaining accuracy in the forensic sciences. *Arizona State Law Journal* 49: 1369–1416.

Koehler, J.J. (2017b). Intuitive error rate estimates for the forensic sciences. *Jurimetrics Journal* 57: 153–168.

Koehler, J.J. (2018). How trial judges should think about forensic science evidence. *Judicature* 102: 28–38.

Koehler, J.J., Chia, A., and Lindsey, J.S. (1995). The random match probability(RMP) in DNA evidence: irrelevant and prejudicial. *Jurimetrics Journal* 35: 201–219.

Kokshoorn, B., Blankers, B.J., de Zoete, J., and Berger, C.E.H. (2017). Activity level DNA evidence evaluation: on propositions addressing the actor or the activity. *Forensic Science International* 278: 115–124.

Köller, N., Nissen, K., Riess, M., and Sadorf, E. (2004). *Probability Conclusions in Expert Opinions on Handwriting. Substantiation and Standardization of Probability Statements in Expert Opinions*. München: Luchterhand.

Korb, K.B. and Nicholson, A.E. (2011). *Bayesian Artificial Intelligence*, 2e. Boca Raton, FL: CRC Press.

Krol, M., Karoly, A., and Koscielniak, P. (2014). Raman spectroscopy and capillary electrophoresis applied to forensic colour inkjet printer inks analysis. *Forensic Science International* 242: 142–149.

Kwan, Q.Y. (1977). Inference of identity of source. PhD thesis. Berkeley: University of California.

Lad, F. (1996). *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*. New York: Wiley.

Lambert, J.A. and Evett, I.W. (1984). The refractive index distribution of control glass samples examined by the forensic science laboratories in the United Kingdom. *Forensic Science International* 26: 1–23.

Lambert, J.A., Satterthwaite, M.J., and Harrison, P.H. (1995a). A survey of glass fragments recovered from clothing of persons suspected of involvement in crime. *Science & Justice* 35: 273–281.

Lambert, J.A., Scranage, J.K., and Evett, I.W. (1995b). Large scale database experiments to assess the significance of matching DNA profiles. *International Journal of Legal Medicine* 108: 8–13.

Lange, K. (1995). Applications of the Dirichlet distributions to forensic match probabilities. In: *Human Identification: The Use of DNA Markers* (ed. B.S. Weir), 107–117. Dordrecht: Kluwer Academic.

Langenburg, G., Hall, C., and Rosemarie, Q. (2015). Utilizing AFIS searching tools to reduce errors in fingerprint casework. *Forensic Science International* 257: 123–133.

Laplace, P.S. (1886). Essai philosophique sur les probabilités. In: *Introduction à la theorie analytique des probabilités; Oeuvres Complètes de Laplace*, vol. 7. Paris: Gauthier-Villars.

Lauritzen, S.L. and Mortera, J. (2002). Bounding the number of contributors to mixed DNA stains. *Forensic Science International* 130: 125–126.

Lee, P.M. (2012). *Bayesian Statistics. An Introduction*, 4e. Chichester: Wiley.

Lee, H.S., Lee, J.W., Han, G.R., and Hwang, J.J. (2000). Motherless case in paternity testing. *Forensic Science International* 114: 57–65.

Lee, J.W., Lee, H.S., and Hwang, J.J. (2002). Statistical analysis for estimating heterogeneity of the Korean population in DNA typing using STR loci. *International Journal of Legal Medicine* 116: 153–160.

Lee, J.W., Lee, H.S., Park, M., and Hwang, J.J. (1999). Paternity probability when a relative of the father is an alleged father. *Science & Justice* 39: 223–230.

Leegwater, A.J., Meuwly, D., Sjerps, M., Vergeer, P., and Alberink, I. (2017). Performance study of a score-based likelihood ratio system for forensic fingermark comparison. *Journal of Forensic Sciences* 62: 626–640.

Lehmann, V.J., Mitchell, R.J., Ballantyne, K.N., and van Oorschot, R.A.H. (2015). Following the transfer of DNA: how does the presence of background DNA affect the transfer and detection of a target source of DNA? *Forensic Science International: Genetics* 19: 68–75.

Lempert, R. (1977). Modelling relevance. *Michigan Law Review* 89: 1021–1057.

Lempert, R. (1986). The new evidence scholarship? Analysing the process of proof. *Boston University Law Review* 66: 439–477.

Lempert, R. (1991). Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. *Cardozo Law Review* 13: 303–341.

Lempert, R. (1993). The suspect population and DNA identification. *Jurimetrics Journal* 34: 1–7.

Lenth, R.V. (1986). On identification by probabilities. *Journal of the Forensic Science Society* 26: 197–213.

Leonard, T. (2000). *A Course in Categorical Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Leonard, T. and Hsu, J.S.J. (1999). *Bayesian Methods*. Cambridge: Cambridge University Press.

Lepot, L. and van den Driessche, T. (2015). Fibre persistence on immersed garment – influence of water flow and stay in running water. *Science & Justice* 55: 431–436.

Lepot, L., van den Driessche, T., Lunstroot, K., Gason, F., and De Wael, K. (2015). Fibre persistence on immersed garment – influence of knitted recipient fabrics. *Science & Justice* 55: 248–253.

Lepot, L., Driessche, T.V., Lunstroot, K., Barret, A., Gason, F., and De Wael, K. (2017). Extraneous fibre traces brought by river water – a case study. *Science & Justice* 57: 53–57.

Levin, E.A., Morgan, R.M., Scott, K.R., and Jones, V.J. (2017). The transfer of diatoms from freshwater to footwear materials: an experimental study assessing transfer, persistence, and extraction methods for forensic reconstruction. *Science & Justice* 57: 349–360.

Levitt, T.S. and Laskey, K.B. (2001). Computational inference for evidential reasoning in support of judicial proof. *Cardozo Law Review* 22: 1691–1731.

Ligertwood, A. and Edmond, G. (2012a). Discussion paper: a just measure of probability. *Law, Probability and Risk* 11: 365–369.

Ligertwood, A. and Edmond, G. (2012b). Expressing evaluative forensic science opinions in a court of law. *Law, Probability and Risk* 11: 289–302.

Lindgren, P., Myrtennäs, K., Forsman, M., Johansson, A., Stenberg, P., Nordgaard, A., and Ahlinder, J. (2019). A likelihood ratio-based approach for improved source attribution in microbiological forensic investigations. *Forensic Science International* 302. https://doi.org/10.1016/j.forsciint.2019.06.027.

Lindley, D.V. (1957). A statistical paradox. *Biometrika* 44: 187–192. Comments by Bartlett, M.S. and Kendall, M.G. appear in *Biometrika* 45: 533–534.

Lindley, D.V. (1975). Probabilities and the law. In: *Utility, Probability, and Human Decision Making* (ed. D. Wendt and C. Vlek), 223–232. Dordrecht: D. Reidel Publishing Company.

Lindley, D.V. (1977a). The distinction between inference and decision. *Synthese* 36: 51–58.

Lindley, D.V. (1977b). Probability and the law. *The Statistician* 26: 203–212.

Lindley, D.V. (1977c). A problem in forensic science. *Biometrika* 64: 207–213.

Lindley, D.V. (1980). L.J. Savage – his work in probability and statistics. *Annals of Statistics* 8: 1–24.
Lindley, D.V. (1985). *Making Decisions*, 2e. London: Wiley.
Lindley, D.V. (1991). Probability. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 27–50. Chichester: Ellis Horwood.
Lindley, D.V. (2000). The philosophy of statistics. *The Statistician* 49: 293–337.
Lindley, D.V. (2004). That wretched prior. *Significance* 1: 85–87.
Lindley, D.V. (2006). *Understanding Uncertainty*. Hoboken, NJ: Wiley.
Lindley, D.V. (2014). *Understanding Uncertainty*, revised edn. Hoboken, NJ: John Wiley & Sons.
Lindley, D.V. and Eggleston, R. (1983). The problem of missing evidence. *Law Quarterly Review* 99: 86–99.
Lindley, D.V. and Scott, W.F. (1995). *New Cambridge Statistical Tables*, 2e. Cambridge: Cambridge University Press.
Locard, E. (1914). La preuve judiciaire par les empreintes digitales. *Archives d?anthropologie criminelle, de médecine légale et de psychologie normale et pathologique* 29: 321–348.
Locard, E. (1920). *L'enquête criminelle et les méthodes scientifiques*. Paris: Flammarion.
Locard, E. (1929). L'analyse des poussières en criminalistique. *Revue Internationale de Criminalistique* 4–5: 176–249.
Locard, E. (1930). The analysis of dust traces. Part I. *The American Journal of Police Science* 1: 276–298.
Locard, E. (1940). *L'enquête criminelle*, Traité de criminalistique, Desvigne, Lyon. Tome septième, Livre VIII.
Louis, T.A. (1981). Confidence intervals for a binomial parameter after observing no successes. *The American Statistician* 35: 154.
Lucas, N., Brown, H., Cook, M., Redman, K., Condon, T., Wrobel, H., Kirkbride, K.P., and Kobus, H. (2016).

A study into the distribution of gunshot residue particles in the random population. *Forensic Science International* 262: 150–155.

Lucena-Molina, J.J. (2016). Epistemology applied to conclusions of expert reports. *Forensic Science International* 264: 122–131.

Lucena-Molina, J.J. (2017). Evaluation of scientific evidence – a proposal on ontological and epistemological bases, and some statistical applications. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Lucena-Molina, J.J., Gascon-Abellan, M., and Pardo-Iranzo, V. (2015a). Technical support for a judge when assessing a prior odds. *Law, Probability and Risk* 14: 147–168.

Lucena-Molina, J.J., Ramos-Castro, D., and Gonzalez-Rodriguez, J. (2015b). Performance of likelihood ratios considering bounds on the probability of observing misleading evidence. *Law, Probability and Risk* 14: 175–192.

Lucy, D. (2005). *Introduction to Statistics for Forensic Scientists*. Chichester: Wiley.

Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley.

Lyon, T.D. and Koehler, J.J. (1996). The relevance ratio: evaluating the probative value of expert testimony in child sexual abuse cases. *Cornell Law Review* 82: 43–78.

Maitre, M., Kirkbride, K.P., Horder, M., Roux, C., and Beavis, A. (2017). Current perspectives in the interpretation of gunshot residues in forensic science: a review. *Forensic Science International* 270: 1–11.

Mallows, C. (1998). The zeroth problem. *The American Statistician* 52: 1–9.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press.

Margiotta, G., Tasselli, G., Tommolini, F., Lancia, M., and Carnevali, E. (2015). Risk of DNA transfer by gloves in forensic casework. *Forensic Science International: Genetics* 5: e527–e529.

Margot, P. (2011). Commentary on 'The need for a research culture in the forensic science'. *UCLA Law Review* 58: 795–801.

Mario, J.R. (2010). A probability-based approach for the analysis of drug seizures composed of multiple containers of either cocaine, heroin or *cannabis*. *Forensic Science International* 197: 105–113.

Marnane, R.N., Elliot, D.A., and Coulson, S.A. (2006). A pilot study to determine the background population of foreign fibre groups on a cotton/polyester T-shirt. *Science & Justice* 46: 215–220.

Marquis, R., Biedermann, A., Cadola, L., Champod, C., Gueissaz, L., Massonnet, G., Mazzella, W.D., Taroni, F., and Hicks, T.N. (2016). Discussion on how to implement a verbal scale in a forensic laboratory: benefits, pitfalls and suggestions to avoid misunderstanding. *Science & Justice* 56: 364–370.

Marquis, R., Taroni, F., Bozza, S., and Schmittbuhl, M. (2006). Quantitative characterization of morphological polymorphism of handwritten characters loops. *Forensic Science International* 164: 211–220.

Marquis, R., Schmittbuhl, M., Mazzella, W.D., and Taroni, F. (2005). Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital letter O. *Forensic Science International* 150: 23–32.

Martire, K.A., Ballantyne, K.N., Bali, A., Edmond, G., Kemp, R.I., and Found, B. (2019). Forensic science evidence: naïve estimates of false positive error rates and reliability. *Forensic Science International* 302. https://doi.org/10.1016/j.forsciint.2019.109877.

Martire, K.A. and Kemp, R.I. (2018). Considerations when designing human performance tests in the forensic sciences. *Australian Journal of Forensic Sciences* 50: 166–182.

Martire, K.A., Kemp, R.I., and Newell, B.R. (2013). The psychology of interpreting expert evaluative opinions. *Australian Journal of Forensic Sciences* 45: 305–314.

Martire, K.A., Kemp, R.I., Sayle, M., and Newell, B.R. (2014). On the interpretation of likelihood ratios in forensic science evidence: presentation formats and the weak evidence effect. *Forensic Science International* 240: 61–68.

Martire, K.A. and Watkins, I. (2015). Perception problems of the verbal scale: a reanalysis and application of a membership function approach. *Science & Justice* 55: 264–273.

Massonnet, G. and Stoecklein, W. (1999). Identification of organic pigments in coatings: applications to red automotive topcoats. Part III: Raman spectroscopy (NIR FT-Raman). *Science & Justice* 39: 181–187.

Matthews, R. (1994). Improving the odds on justice. *New Scientist* 142 (1921): 12.

Mavridis, D. and Aitken, C.G.G. (2009). Sample size determination for categorical responses. *Journal of Forensic Sciences* 54: 135–151.

Maxwell, J.C. (1990). Texts. In: *The Scientific Letters and Papers of James Clerk Maxwell, 1846–1862*, Vol. 1 (ed. J.C. Maxwell and P.M. Harman), 31–859. Cambridge: Cambridge University Press.

Mazzella, W.D. and Marquis, R. (2007). Forensic image analysis of laser-printed documents. *Journal of the American Society of Questioned Documents* 10: 19–24.

McDermott, Y. and Aitken, C.G.G. (2017). Analysis of evidence in international criminal trials using Bayesian belief networks. *Law, Probability and Risk* 16: 111–129.

McDermott, S.D., Willis, S.M., and McCullough, J.P. (1999). The evidential value of paint. Part II: A Bayesian approach. *Journal of Forensic Sciences* 44: 263–269.

McKenna, L. (2013). Understanding DNA results within the case context: importance of the alternative proposition. *Frontiers in Genetics* 4: 1–3.

McQuillan, J. and Edgar, K. (1992). A survey of the distribution of glass on clothing. *Journal of the Forensic Science Society* 32: 333–348.

Meakin, G.E., Butcher, E.V., and van Oorschot, R.A.H. (2017). Trace DNA evidence dynamics: an investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics* 29: 38–47.

Meakin, G.E. and Jamieson, A. (2013). DNA transfer: review and implications for casework. *Forensic Science International: Genetics* 7: 434–443.

Meakin, G.E. and Jamieson, A. (2016). A response to a response to Meakin and Jamieson (2013). *Forensic Science International: Genetics* 22: e5–e6.

Meester, R., Collins, M., Gill, R., and van Lambalgen, M. (2006). On the (ab)use of statistics in the legal case against the nurse Lucia de B. *Law, Probability and Risk* 5: 233–250.

Meester, R. and Sjerps, M. (2003). The evidential value in the DNA database search controversy and the two-stain problem. *Biometrics* 59: 727–732.

Meester, R. and Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. *Law, Probability and Risk* 3: 51–62.

Meester, R. and Slooten, K. (2019a). Calculating the posterior odds from a single-match DNA database search under various scenarios with minimal assumptions. *Law, Probability and Risk* 18: 223–228.

Meester, R. and Slooten, K. (2019b). Ne bis in idem – a commentary on 'Calculating posterior odds from a single-match DNA database search'. *Law, Probability and Risk* 18: 35–38.

Meier, P. and Zabell, S. (1980). Benjamin Peirce and the Howland will. *Journal of the American Statistical Association* 75: 497–506.

Mellen, B.G. (2000). A likelihood approach to DNA evidence. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 125–141. New York: Springer-Verlag.

Mellen, B.G. and Royall, R. (1997). Measuring the strength of deoxyribonucleic acid evidence and probabilities of strong implicating evidence. *Journal of the Royal Statistical Society, Series A* 160: 305–320.

Meuwly, D. (2001). Reconnaissance de locuteurs en sciences forensiques: l'apport d'une approche automatique. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Meuwly, D. and Drygajlo, A. (2001). Forensic speaker recognition based on a Bayesian framework and Gaussian mixture modelling (GMM). Proceedings of the 2001 Speaker Odyssey Recognition Workshop, Crete, Greece (18–22 June 2000), pp. 145–150.

Meuwly, D., Ramos-Castro, D., and Haraksim, R. (2017). A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Science International* 276: 142–153.

Miles, R.F. (2007). The emergence of decision analysis. In: *Advances in Decision Analysis* (ed. W. Edwards, R.F. Miles, and D. von Winterfelt), 13–31. Cambridge: Cambridge University Press.

Miller, L.S. (1987). Procedural bias in forensic science examinations of human hair. *Law and Human Behaviour* 11: 157–163.

Milot, E., Baechler, S., and Crispino, F. (2020). Must the random man be unrelated? A lingering misconception in forensic genetics. *Forensic Science International: Synergy* 2: 35–40.

Mnookin, J.L. (2008). The validity of latent fingerprint identification: confessions of a fingerprinting moderate. *Law, Probability and Risk* 7: 127–141.

Mode, E.B. (1963). Probability and criminalistics. *Journal of the American Statistical Association* 58: 628–640.

Monari, P. and Cocchi, D. (eds.) (1993). Probabilità e induzione, 105–114. Bologna: Clueb.

Moore, R., Kingsbury, D., Bunford, J., and Tucker, V.C. (2012). A survey of paint flakes on the clothing of persons suspected of involvement in crime. *Science & Justice* 52: 96–101.

Moras, C. (1906). L'Affaire Dreyfus: les débuts de la Cour de Cassation (15 Juin 1906 – 12 Juillet 1906). Paris: Société Nouvelle de Librairie et d'édition.

Moretti, T.R. and Budowle, B. (2017). Letter to the editor – reiteration of the statistical basis of DNA source attribution determinations in view of the Attorney General's Directive on 'Reasonable Scientific Certainty' statements. *Journal of Forensic Sciences* 62 (4): 1114–1115.

Moretti, T.R., Just, R.S., Kehl, S.C., Willis, L.E., Buckleton, J.S., Bright, J.A., Taylor, D., and Onorato, A.J. (2017). Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 29: 126–144.

Morgan, R.M., Allen, E., King, T., and Bull, P.A. (2014). The spatial and temporal distribution of pollen in a room: forensic implications. *Science & Justice* 54: 49–56.

Morgan, J.P., Chaganty, N.R., Dahiya, R.C., and Doviak, M.J. (1991). Let's make a deal: the player's dilemma. *The American Statistician* 45 284–287.

Moroni, R., Aalberg, L., Reinikainen, T., and Corander, J. (2012). Bayesian adaptive approach to estimating sample sizes for seizures of illicit drugs. *Journal of Forensic sciences* 57: 80–85.

Morris, K.B., Law, E.F., Jeffreys, R.L., Dearth, E.C., and Fabyanic, E.B. (2017). An evaluation of the discriminating power of an integrated ballistics identification system heritage system with the NIST standard cartridge case (Standard Reference Material 2461). *Forensic Science International* 280: 188–193.

Morrison, G.S. (2011). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic phonetic data multivariate kernel density (MKVD)

versus Gaussian mixture model-universal background model (GMM-UBM). *Speech Communication* 53: 242–256.

Morrison, G.S. and Enzinger, E. (2016). What should a forensic practitioner's likelihood ratio be? *Science & Justice* 56: 374–379.

Morrison, G.S. and Enzinger, E. (2018). Score-based procedures for the calculation of forensic likelihood ratios – scores should take account of both similarity and typicality. *Science & Justice* 58: 47–58.

Morrison, G.S., Sahito, F.H., Jardine, G., Djokic, D., Clavet, S., Berghs, S., and Goemans Dorny, C. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science international* 263: 92–100.

Mortera, J., Dawid, A.P., and Lauritzen, S.L. (2003). Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology* 63: 191–205.

Mortera, J., Vecchiotti, C., Zoppis, S., and Merigioli, S. (2016). Paternity testing that involves a DNA mixture. *Forensic Science International: Genetics* 23: 50–54.

Muehlethaler, C., Massonnet, G., and Esseiva, P. (2014). Discrimination and classification of FTIR spectra of red, blue and green spray paints using a multivariate statistical approach. *Forensic Science International* 244: 170–178.

Mullen, C., Spence, D., Moxey, L., and Jamieson, A. (2014). Perception problems of the verbal scale. *Science & Justice* 54: 154–158.

Murrie, D.C., Gardner, B.O., Kelley, S., and Dror, I.E. (2019). Perceptions and estimates of error rates in forensic science: a survey of forensic analysts. *Forensic Science International* 302. https://doi.org/10.1016/j.forsciint.2019.109887.

Nance, D.A. (2019). Belief functions and burdens of proof. *Law, Probability and Risk* 18: 53–76.

National Research Council (1996). *The Evaluation of Forensic DNA Evidence*, 2e. Washington, DC: National Academy Press.

Neil, M., Fenton, N., Lagnado, D., and Gill, R. (2019). Modelling competing legal arguments using Bayesian model comparison and averaging. *Artificial Intelligence and Law* 27: 403–430.

Neocleous, T., Aitken, C.G.G., and Zadora, G. (2011). Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometrics and Intelligent Laboratory Systems* 109: 77–85.

Neumann, C. and Ausdemore, M.A. (2019). Communicating forensic evidence: is it appropriate to report posterior beliefs when DNA evidence is obtained through a database search? *Law, Probability and Risk* 18: 25–34.

Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Meuwly, D., and Bromage-Griffiths, A. (2006). Computation of likelihood ratios in fingerprint identification for configurations of three minutiae. *Journal of Forensic Sciences* 51: 1255–1266.

Neumann, C. and Saunders, C.P. (2015). Commentary on: Alberink I, de Jongh A, Rodriguez C, Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. *Journal of Forensic Sciences* 2014 59 (1): 70–81; *Journal of Forensic Sciences* 60: 252–256.

Neumann, C. and Stern, H.S. (2016). Forensic examination of fingerprints: past, present, and future. *Chance* 29: 9–16.

Noël, S., Lagacé, K., Rogic, A., Granger, D., Bourgoin, S., Jolicoeur, C., and Séguin, D. (2016). DNA transfer during laundering may yield complete genetic profiles. *Forensic Science International: Genetics* 23: 240–247.

Nordgaard, A. (2016). Comment on Taroni et al. (2016a). *Law, Probability and Risk* 15: 17–22.

Nordgaard, A., Ansell, R., Drotz, W., and Jaeger, L. (2012a). Scale of conclusions for the value of evidence. *Law, Probability and Risk* 11: 1–24.

Nordgaard, A., Hedell, R., and Ansell, R. (2012b). Assessment of forensic findings when alternative explanations have different likelihoods – blame-the-brother syndrome. *Science & Justice* 52: 226–236.

Nordgaard, A. and Rasmusson, B. (2012a). The likelihood ratio as value of evidence – more than a question of numbers. *Law, Probability and Risk* 11: 303–315.

Nordgaard, A. and Rasmusson, B. (2012b). A short reply on the discussion by D. Hamer. *Law, Probability and Risk* 11: 371.

O'Hagan, A. (2019). Expert knowledge elicitation: subjective but scientific. *The American Statistician* 73: 69–81.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., and Rakow, T. (2006). *Uncertain Judgements. Eliciting Expert's Probabilities*. Chichester: Wiley.

Olkin, I. (1958). The evaluation of physical evidence and the identity problem by means of statistical probabilities. *General Session of the American Academy of Forensic Sciences*, Cleveland, Ohio, U.S.A.

Ommen, D.M. and Saunders, C.P. (2018). Building a unified statistical framework for the forensic identification of source problems. *Law, Probability and Risk* 17: 179–197.

Ommen, D.M., Saunders, C.P., and Neumann, C. (2016). An argument against presenting interval quantifications as a surrogate for the value of evidence. *Science & Justice* 56: 383–387.

Ommen, D.M., Saunders, C.P., and Neumann, C. (2017). The characterization of Monte Carlo errors for the quantification of the value of forensic evidence. *Journal of Statistical Computation and Simulation* 87: 1608–1643.

Ostrum, R.B. (2019). CSFS document section position on the logical approach to evidence evaluation and

corresponding wording of conclusions. *Canadian Society of Forensic Science Journal* 52: 129–138.

O'Sullivan, S., Geddes, T., and Lovelock, T.J. (2011). The migration of fragments of glass from the pockets to the surfaces of clothing. *Forensic Science International* 208: 149–155.

Page, H., Sarna, A., Watts, L., Ward, E., and McKenzie, M. (2014). The recovery of semen from bathwater using the evidence recovery system (ERS). *Science & Justice* 54: 89–94.

Page, M., Taylor, J., and Blenkin, M. (2011). Uniqueness in the forensic identification sciences – fact or fiction? *Forensic Science International* 206: 12–18.

Palmer, R. (2016). The evaluation of fibre evidence in the investigation of serious crime. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Palmer, R. and Banks, M. (2005). The secondary transfer of fibres from head hair. *Science & Justice* 45: 123–128.

Palmer, R. and Burch, H.J. (2009). The population, transfer and persistence of fibres on the skin of living subjects. *Science & Justice* 49: 259–264.

Palmer, R. and Chinherende, V. (1996). A target fibre study using cinema and car seats as recipient item. *Journal of the Forensic Science Society* 41: 802–803.

Palmer, R. and Oliver, S. (2004). The population of coloured fibres in human head hair. *Science & Justice* 44: 83–88.

Palmer, R. and Polwarth, G. (2011). The persistence of fibres on skin in an outdoor deposition crime scene scenario. *Science & Justice* 51: 187–189.

Palmer, R., Sheridan, K., Puckett, J., Richardson, N., and Lo, W. (2017). An investigation into secondary transfer – the transfer of textile fibres. *Forensic Science International* 278: 334–337.

Parker, J.B. (1966). A statistical treatment of identification problems. *Forensic Science Society Journal* 6: 33–39.

Parker, J.B. (1967). The mathematical evaluation of numerical evidence. *Forensic Science Society Journal* 7: 134–144.

Parker, J.B. and Holford, A. (1968). Optimum test statistics with particular reference to a forensic science problem. *Applied Statistics* 17: 237–251.

Parmigiani, G. (2002). *Modeling in Medical Decision Making*. Chichester: Wiley.

Peabody, A.J., Oxborough, R.J., Cage, P.E., and Evett, I.W. (1983). The discrimination of cat and dog hairs. *Journal of the Forensic Science Society* 23: 121–129.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers.

Pearson, E.S. and Kendall, M.G. (1970). *Studies in the History of Statistics and Probability*. London: Charles Griffin.

Pearson, E.F., May, R.W., and Dabbs, M.G.D. (1971). Glass and paint fragments found in men's outer clothing – report of a survey. *Journal of Forensic Sciences* 16: 283–300.

Pedroso, J.F., Ramos-Castro, D., and Gonzalez-Rodriguez, J. (2016). Gaussian mixture models of between-source variation for likelihood ratio computation from multivariate data. *Plos ONE* 11 (2): 1–25.

Peirce, C.S. (1878). Probability. Reprinted in: *Writings of Charles S. Peirce (1986)* (ed. C.J.W. Kloesel), 290–305. Bloomington, IN: Indiana University Press.

Petraco, N.D.K., Shenkin, P., Speir, J., Diaczuk, P., Pizzola, P., Gambino, C., and Petraco, N. (2012). Addressing the National Academy of Sciences challenge: a method for statistical pattern comparison of striated tool marks. *Journal of Forensic Sciences* 57: 900–911.

Piattelli-Palmarini, M. (1994). *Inevitable Illusions*. New York: Wiley.

Pighin, S., Girotto, V., and Tentori, K. (2017). Children's quantitative Bayesian inferences from natural frequencies and number of chances. *Cognition* 168: 164–175.

Pittella, J.E.H. and Gusmäo, S.N.S. (2003). Diffuse vascular injury in fatal road traffic accident victims:

its relationship to diffuse axonal injury. *Journal of Forensic Sciences* 48: 626–630.

Poincaré, H. (1896). Calcul des probabilitiés: leçons professées pendant le deuxième semestre 1893–1894. *Calcul des probabilitiés*, Carré, G. Paris.

Poincaré, H. (1912). *Calcul des probabilités*. Paris: Gauthier-Villars.

Poincaré, H. (1992). *La Science et l'Hypothèse*. Paris: Editions de la Bohème.

Poisson, S.D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matiere civile. précédées des règles generales du calcul des probabilités*. Paris: Bachelier.

Pounds, C.A. and Smalldon, K.W. (1978). The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby. *Journal of the Forensic Science Society* 18: 197–203.

President's Council of Advisors on Science and Technology (PCAST) (2016). Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods.

Press, S.J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2e. Malibar: Robert E. Krieger Publishing Company.

Press, S.J. (1989). *Bayesian Statistics: Principles, Models and Applications*. New York: Wiley.

Press, S.J. (2003). *Subjective and Objective Bayesian Statistics*. New York: Wiley.

Press, S.J. and Tanur, J.M. (2001). *The Subjectivity of Scientists and the Bayesian Approach*. Mineola, NY: Dover Publications.

Puch-Solis, R., Roberts, P., Pope, S., and Aitken, C.G.G. (2013). *Assessing the Probative Value of DNA Evidence* (Practitioner Guide No. 2). Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, London: Royal Statistical Society.

Puch-Solis, R. and Smith, J.Q. (2002). FINDS: a training package to assess forensic fibre evidence. In: *MICAI 2002: Advances in Artificial Intelligence* (ed. C.A. Coello Coello, A. de Albornoz, L.E. Sucar, and O.C. Battistutti), 420–429. Berlin: Springer-Verlag.

Pun, K.M. (2016). Interprétations des profils génétiques obtenus à partir de traces de contact. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Rabinovitch, N.L. (1969). Studies in the history of probability and statistics XXII: probability in the Talmud. *Biometrika* 56: 437–441.

Rabinovitch, N.L. (1973). *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. Toronto: University of Toronto Press.

Race, R.R., Sanger, R., Lawler, S.D., and Bertinshaw, D. (1949). The inheritance of the MNS blood groups: a second series of families. *Heredity* 3: 205–213.

Rahne, E., Joseph, L., and Gyorkos, T.W. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Applied Statistics* 49: 119–128.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Cambridge, MA: MIT Press.

Ramos-Castro, D. and Gonzalez-Rodriguez, J. (2013). Reliable support: measuring calibration of likelihood ratios. *Forensic Science International* 230: 156–169.

Ramos-Castro, D., Gonzalez-Rodriguez, J., Zadora, G., and Aitken, C.G.G. (2013). Information-theoretical assessment of the performance of likelihood ratio computation methods. *Journal of Forensic Sciences* 58: 1503–1518.

Ramos-Castro, D., Krish, R.P., Fierrez, J., and Meuwly, D. (2017). From biometric scores to forensic likelihood ratios. In: *Handbook of Biometrics for Forensic Science, Advances in Computer Vision and Pattern Recognition* (ed. C. Champod and M. Tistarelli), 305–327. Cham: Springer.

Ramsey, F.P. (1931). Truth and probability. In: *The Foundations of Mathematics and Other Logical Essays* (ed. R.B. Braithwaite). 156–198 London: Routledge & Kegan Paul. https://core.ac.uk/download/pdf/7048428.pdf.

Redmayne, M. (1995). Doubts and burdens: DNA evidence, probability and the courts. *Criminal Law Review* 464–482.

Redmayne, M.. (1997). Presenting probabilities in court: the DNA experience. *The International Journal of Evidence and Proof* 4: 187–214.

Redmayne, M. (2002). Appeals to reason. *The Modern Law Review* 65: 19–35.

Reed, G., Lofts, C., and Coyle, T. (2010). A population study of polyurethane foam fragments recovered from the surface of 100 outer-garments. *Science & Justice* 50: 127–137.

Reinstein, R.S. (1996). Comment. In: *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial* (ed. E. Connors, T. Lundregan, N. Miller, and T. McEwen). Washington, DC: US Department of Justice. xxi–xxii. https://www.ncjrs.gov/pdffiles/dnaevid.pdf.

Ribeiro, G., Tangen, J.M., and McKimmie, B.M. (2019). Beliefs about error rates and human judgment in forensic science. *Forensic Science International* 297: 138–147.

Riva, F. (2011). Etude de la valeur indicielle des traces présentes sur les douilles. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Riva, F. and Champod, C. (2014). Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences* 59: 637–647.

Riva, F., Hermsen, R., Mattijssen, E., Pieper, P., and Champod, C. (2017). Objective evaluation of subclass characteristics on breech face marks. *Journal of Forensic Sciences* 62: 417–422.

Robert, C. (2007). *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*, 2e. New York: Springer-Verlag.

Robert, C. (2013). Des spécificités de l'approche Bayesiénne et de ses justifications en statistique inférentielle. *HAL*. Available at https://hal.archives-ouvertes.fr. Id: hal.00870124.

Roberts, P. and Aitken, C.G.G. (2013). *The Logic of Forensic Proof* (Practitioner Guide No. 3). Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, London: Royal Statistical Society.

Robertson, B. and Vignaux, G.A. (1998). Explaining evidence logically. *New Law Journal* 148: 159–162.

Robertson, B. and Vignaux, G.A. (1992). Unhelpful evidence in paternity cases. *New Zealand Law Journal* 9: 315–317.

Robertson, B. and Vignaux, G.A. (1993a). Biology, logic, statistics and criminal justice. *The Criminal Lawyer* 35: 1–6.

Robertson, B. and Vignaux, G.A. (1993b). Probability – the logic of the law. *Oxford Journal of Legal Studies* 13: 457–478.

Robertson, B. and Vignaux, G.A. (1994). Crime investigation and the criminal trial (Lether to the Editor). *Journal of the Forensic Science Society* 34: 270.

Robertson, B. and Vignaux, G.A. (1995a). DNA evidence: wrong answers or wrong questions? In: *Human Identification: The Use of DNA Markers* (ed. B.S. Weir), 145–152. Dordrecht: Kluwer Academic.

Robertson, B. and Vignaux, G.A. (1995b). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.

Robertson, B., Vignaux, G.A., and Berger, C.E.H. (2016). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, 2e. Chichester: Wiley.

Robinson, N., Taroni, F., Saugy, M., Ayotte, C., Mangin, P., and Dvorak, J. (2001). Detection of nandrolone

metabolites in urine after a football game in professional and amateur football players: a Bayesian comparison. *Forensic Science International* 122: 130–135.

Roewer, L. and Geppert, M. (2012). Interpretation guidelines of a standard Y-chromosome STR 17-plex PCR-CE assay for crime casework. In: *DNA Electrophoresis Protocols for Forensic Genetics* (ed. A. Alonso), 43–56. New York: Springer-Verlag.

Rouder, J.N., Wixted, J.T., and Christenfeld, N.J.S. (2019). Rejoinder for calculating the posterior odds from a single-match DNA database search. *Law, Probability and Risk* 18: 43–51.

Roux, C., Langdon, S., Waight, D., and Robertson, J. (1999). The transfer and persistence of automotive carpet fibres on shoe soles. *Science & Justice* 39: 239–251.

Roux, C. and Margot, P. (1997). An attempt to assess relevance of textile fibres recovered from car seats. *Science & Justice* 37: 225–230.

Roux, C. and Wiggins, K.G. (2017). Aids to interpretation. In: *Forensic Examination of Fibres* (ed. J. Robertson and C. Roux), 375–394. London: Taylor & Francis.

Royal Society and Royal Society of Edinburgh (2017). Forensic DNA analysis: a primer for courts, London and Edinburgh. https://royalsociety.org/~/media/about-us/programmes/science-and-law (accessed 09 June 2020).

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Royall, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association* 95: 760–768.

Ryan, B.F., Joiner, B.L., and Ryan, T.A. (2000). *MINITAB Handbook*, 4e. Pacific Grove, CA: Brooks Cole. http://www.minitab.com/.

Sacco, N.R. (1969). *The Sacco–Vanzetti Case: Transcript of the Record of the Trial of Nicola Sacco and Bartolomeo*

Vanzetti in the Courts of Massachusetts and Subsequent Proceedings, 1920–1927. Mamaroneck, NY: P. P. Appel.

Saks, M.J. and Koehler, J.J. (1991). What DNA 'fingerprinting' can teach the law about the rest of forensic science. *Cardozo Law Review* 13: 361–372.

Saks, M.J. and Koehler, J.J. (2005). The coming paradigm shift in forensic identification science. *Science* 309 (5736): 892–895.

Saks, M.J. and Koehler, J.J. (2008). The individualization fallacy in forensic science evidence. *Vanderbilt Law Review* 61: 199–219.

Salmon, D. and Salmon, C. (1980). Blood groups and genetic markers polymorphisms and probability of paternity. *Transfusion* 20: 684–694.

Salter, M.T. and Cook, R. (1996). Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International* 81: 211–221.

Samie, L. (2019). Evaluation des résultats ADN considérant des propositions au niveau de l'activité. PhD thesis. Lausanne: The University of Lausanne, School of Criminal Justice.

Samie, L., Champod, C., Glutz, V., Garcia, M., Castella, V., and Taroni, F. (2019). The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry. *Science & Justice* 59: 405–410.

Samie, L., Hicks, T.N., Castella, V., and Taroni, F. (2016). Stabbing simulations and DNA transfer. *Forensic Science International: Genetics* 22: 73–80.

Samie, L., Taroni, F., and Champod, C. Estimating the quantity of transferred DNA in primary and secondary transfers. (2020). *Science & Justice*. https://doi.org/10.1016/j.scijus.2019.09.008.

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Dover Publications.

Savage, L.J. (1967). Difficulties in the theory of personal probability. *Philosophy of Science* 34: 305–310.

Schield, C., Campelli, C., Sycalik, J., Randle, C., Hughes-Stamm, S., and Gangitano, D. (2016).

Identification and persistence of Pinus pollen DNA on cotton fabrics: a forensic application. *Science & Justice* 56: 29–34.

Schneider, P.M., Schneider, H., Fimmers, R., Keil, K., Molsberger, G., Pflug, W., Rothämel, T., Eckert, M., Pfeiffer, H., and Brinkmann, B. (2010). Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNA-Datenbank-Treffern. *Rechtsmedizin* 20: 111–115.

Schrödinger, E. (1947). The foundation of the theory pf probability – I. *Proceedings of the Royal Irish Academy – Section A: Mathematical and Physical Sciences* 51: 51–66.

Schum, D.A. (1994). *Evidential Foundations of Probabilistic Reasoning*. Chichester: Wiley.

Schum, D.A. (1999). Inference networks and the evaluation of evidence: alternative analyses. In: *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference (UAI-1999)* (ed. K. Laskey and H. Prade), 575–584. San Francisco, CA: Morgan Kaufmann Publishers.

Schum, D.A. (2001). *Evidential Foundations of Probabilistic Reasoning*. Evanston, IL: Northwestern University Press.

Schum, D.A. (2009). A science of evidence: contributions from law and probability. *Law, Probability and Risk* 8: 197–231.

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2009). Y-chromosome Short Tandem Repeat (Y-STR) interpretation guidelines. *Forensic Science Communications*. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2009/standards/2009_01_standards01.htm.

Scott, D.W. (1992). *Multivariate Density Estimation*. New York: Wiley.

Scozzafava, R. (1987). Subjective probability and Bayesian statistics in engineering mathematics education. *International Journal of Mathematical Education in Science and Technology* 18: 685–688.

Seheult, A. (1978). On a problem in forensic science. *Biometrika* 65: 646–648.
Selvin, S. (1975). On the Monty Hall problem. *The American Statistician* 29: 134.
Shachter, R.D. (1986). Evaluating influence diagrams. *Operations Research* 34: 871–882.
Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.
Shafer, G. (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences* 19: 309–370.
Shafer, G. (1982). Lindley's paradox (with discussion). *Journal of the American Statistical Association* 77: 325–351.
Shafer, G. (2019). On the nineteenth-century origins of significance testing and p-hacking. https://bit.ly/2p8BSBP (last accessed 23 December 2019).
Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379–243.
Sheynin, O.B. (1974). On the prehistory of the theory of probability. *Archive for History of Exact Sciences* 12: 97–141.
Shoemaker, J.S., Painter, I.S., and Weir, B.S. (1999). Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics* 15: 354–358.
Siegel, J.A. (1997). Evidential value of textile fiber – transfer and persistence of fibers. *Forensic Science Review* 9: 81–96.
Silverman, B.W. (1986). *Density Estimation*. London: Chapman & Hall.
Simons, A.A. (1997). Technical working group on friction ridge analysis, study and technology (TWGFAST) guidelines. *Journal of Forensic Identification* 48: 147–162.
Simpson, E.H. (1949). Measures of diversity. *Nature* 163: 688.

Sironi, E., Pinchi, V., Pradella, F., Focardi, M., Bozza, S., and Taroni, F. (2018). Bayesian networks of age estimation and classification based on dental evidence: a study on the third molar mineralization. *Journal of Forensic and Legal Medicine* 55: 23–32.

Sjerps, M. (2019). Reporting DNA database matches: we need more research. *Law, Probability and Risk* 18: 39–41.

Sjerps, M., Alberink, I., Bolck, A., Stoel, R.D., Vergeer, P., and van Zanten, J.H. (2016). Uncertainty and LR: to integrate or not to integrate: that's the question. *Law, Probability and Risk* 15: 23–29.

Sjerps, M. and Berger, C.E.H. (2012). How clear is transparent? Reporting expert reasoning in legal cases. *Law, Probability and Risk* 11: 317–329.

Sjerps, M. and Kloosterman, A.D. (1999). On the consequences of DNA profile mismatches for close relatives of an excluded person. *International Journal of Legal Medicine* 112: 176–180.

Slooten, K. and Ricciardi, F. (2013). Estimation of mutation probabilities for autosomal STR markers. *Forensic Science International: Genetics* 7: 337–344.

Smalldon, K.W. and Moffat, A.C. (1973). The calculation of discriminating power for a series of correlated attributes. *Journal of the Forensic Science Society* 13: 291–295.

Smeeton, N.C. and Adcock, C.J. (1997). Foreword. *The Statistician* 46 (2): 127.

Smith, R.L. and Charrow, R.P. (1975). Upper and lower bounds for the probability of guilt based on circumstantial evidence. *Journal of the American Statistical Association* 70: 555–560.

Song, J., Vorburger, T.V., Chu, W., Yen, J., Soons, J.A., Ott, D.B., and Zhang, N.F. (2018). Estimating error rates for firearm evidence identifications in forensic science. *Forensic Science International* 284: 15–32.

Souder, W. (1934). The merits of scientific evidence. *Journal of the American Institute of Criminal Law and Criminology* 25: 683–684.

Steensma, K., Ansell, R., Clarisse, L., Connolly, E., Kloosterman, A.D., McKenna, L., van Oorschot, R.A.H., Szkuta, B., and Kokshoorn, B. (2017). An inter-laboratory comparison study on transfer, persistence and recovery of DNA from cable ties. *Forensic Science International: Genetics* 31: 95–104.

Stockmarr, A. (1999). Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics* 55: 671–677.

Stockton, A. and Day, S. (2001). Bayes, handwriting and science. *Proceedings of the 59th Annual ASQDE Meeting – Handwriting & Technology: at the Crossroads*, Des Moines, Iowa, USA, pp. 1–10.

Stoney, D.A. (1984). Evaluation of associative evidence: choosing the relevant question. *Journal of the Forensic Science Society* 24: 473–482.

Stoney, D.A. (1985). *Quantitative assessment of fingerprint individuality*. PhD thesis. Berkeley, CA: The University of California at Berkeley.

Stoney, D.A. (1991a). Transfer evidence. In: *The Use of Statistics in Forensic Science* (ed. C.G.G. Aitken and D.A. Stoney), 107–138. Ellis Horwood.

Stoney, D.A. (1991b). What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society* 31: 197–199.

Stoney, D.A. (1992). Reporting of highly individual genetic typing results: a practical approach. *Journal of Forensic Sciences* 37: 373–386.

Stoney, D.A. (1994). Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *Journal of the Forensic Science Society* 34: 17–21.

Stoney, D.A. (2012) Discussion on the paper by Neumann, C., Evett, I.W., and Skerrett, J. (2012). Quantifying the weight of evidence from a fingerprint

comparison: a new paradigm. *Journal of the Royal Statistical Society, Series A* 175: 371–415; *Journal of the Royal Statistical Society, Series A* 175: 399–400.

Stoney, D.A. and Thornton, J.I. (1986). A critical analysis of quantitative fingerprint individuality models. *Journal of Forensic Sciences* 31: 1187–1216.

Stoney, D.A. and Thornton, J.I. (1988). Authors' response for discussion of Stoney and Thornton (1986). *Journal of Forensic Sciences* 33: 11–12.

Swofford, H.J., Koertner, A.J., Zemp, F., Ausdemore, M.A., Liu, A., and Salyards, M.J. (2018). A method for the statistical interpretation of friction ridge skin impression evidence: method development and validation. *Forensic Science international* 287: 113–126.

Szkuta, B., Ballantyne, K.N., Kokshoorn, B., and van Oorschot, R.A.H. (2018). Transfer and persistence of non-self DNA on hands over time: using empirical data to evaluate DNA evidence given activity level propositions. *Forensic Science International: Genetics* 33: 84–97.

Szkuta, B., Harvey, M.L., Ballantyne, K.N., and van Oorschot, R.A.H. (2015). DNA transfer by examination tools – a risk for forensic casework? *Forensic Science International: Genetics* 16: 246–254.

Taroni, F. (2018). DNA evidence at trial: experiences as expert witness for the defence. *3rd International Symposium on Sino-Swiss Evidence Science*, Hangzhou City, (25–27 June 2018).

Taroni, F. and Aitken, C.G.G. (1998a). Probabilités et preuve par l'ADN dans les affaires civiles et criminelles. Questions de la cour et réponses fallacieuses des experts. *Revue Pénale Suisse* 116: 291–313.

Taroni, F. and Aitken, C.G.G. (1998b). Probability reasoning in the law, Part 1: Assessment of probabilities and explanation of the value of DNA evidence. *Science & Justice* 38: 165–177.

Taroni, F. and Aitken, C.G.G. (1998c). Probability reasoning in the law, Part 2: Assessment of probabilities and

explanation of the value of trace evidence other than DNA. *Science & Justice* 38: 179–188.

Taroni, F. and Aitken, C.G.G. (1999). The likelihood approach to compare populations: a study on DNA evidence and pitfalls of intuition. *Science & Justice* 39: 213–222.

Taroni, F. and Aitken, C.G.G. (2000). Letter to the editor: fibres evidence, probabilistic evaluation and collaborative test. *Forensic Science International* 114: 45–47.

Taroni, F., Aitken, C.G.G., and Garbolino, P. (2001). De Finetti's subjectivism, the assessment of probabilities, and the evaluation of evidence: a commentary for forensic scientists. *Science & Justice* 41: 145–150.

Taroni, F. and Biedermann, A. (2005). Inadequacies of posterior probabilities for the assessment of scientific evidence. *Law, Probability and Risk* 4: 89–114.

Taroni, F., Biedermann, A., and Bozza, S. (2016). Statistical hypothesis testing and common misinterpretations: should we abandon p-value in forensic science applications? *Forensic Science International* 259: e32–e36.

Taroni, F., Biedermann, A., Bozza, S., Comte, J., and Garbolino, P. (2012a). Uncertainty about the true source – a note on the likelihood ratio at the activity level. *Forensic Science International* 220: 173–179.

Taroni, F., Marquis, R., Schmittbuhl, M., Biedermann, A., Thiéry, A., and Bozza, S. (2012b). The use of the likelihood ratio for evaluative and investigative purposes in comparative forensic handwriting examination. *Forensic Science International* 214: 189–194.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., and Aitken, C.G.G. (2014a). *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*, 2e. Chichester: Wiley.

Taroni, F., Biedermann, A., Bozza, S., Vuille, J., and Augsburger, M. (2014b). Toxic substances in blood: an analysis of current recommendations under a Bayesian (decision) approach. *Law, Probability and Risk* 13: 27–45.

Taroni, F., Marquis, R., Schmittbuhl, M., and Bozza, S. (2014c). Bayes factor for investigative assessment of selected handwriting features. *Forensic Science International* 242: 266–273.

Taroni, F., Biedermann, A., Coquoz, R., Hicks, T.N., and Champod, C. (2011). Brief an den Herausgeber zum Beitrag von Schneider, et al., Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNA-Datenbank-Treffern (Letter to the Editor with reference to Schneider et al. (2010). Recommendations of the German Stain Commission regarding the statistical evaluation of matches following searches in the national DNA database). *Rechtsmedizin* 21: 55–57.

Taroni, F., Biedermann, A., Garbolino, P., and Aitken, C.G.G. (2004). A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International* 139: 5–16.

Taroni, F., Biedermann, A., Vuille, J., and Morling, N. (2013). Whose DNA is this? This is not the relevant question (a note for forensic scientists). *Forensic Science International: Genetics* 7: 467–470.

Taroni, F., Bozza, S., and Aitken, C.G.G. (2005). Decision analysis in forensic science. *Journal of Forensic Sciences* 50: 894–905.

Taroni, F., Bozza, S., Bernard, M., and Champod, C. (2007). Value of DNA tests: a decision perspective. *Journal of Forensic Sciences* 52: 31–39.

Taroni, F., Bozza, S., and Biedermann, A. (2006). Two items of evidence, no putative source: an inference problem in forensic intelligence. *Journal of Forensic Sciences* 51: 1350–1361.

Taroni, F., Bozza, S., and Biedermann, A. (2020). Decision theory. In: *Handbook of Forensic Statistics* (ed. D. Banks, K. Kafadar, D. Kaye, and M. Tackett). Boca Raton, FL: CRC Press. 103–130.

Taroni, F., Bozza, S., Biedermann, A., and Aitken, C.G.G. (2016). Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. *Law, Probability and Risk* 15: 1–16.

Taroni, F., Bozza, S., Biedermann, A., Garbolino, P., and Aitken, C.G.G. (2010). *Data Analysis in Forensic Science. A Bayesian Decision Perspective*. Chichester: Wiley.

Taroni, F., Bozza, S., Hicks, T.N., and Garbolino, P. (2019). More on the question 'when does absence of evidence constitute evidence of absence?' How Bayesian confirmation theory can logically support the answer. *Forensic Science International* 301: e59–e63.

Taroni, F. and Buckleton, J.S. (2002). Likelihood ratio as a relevant and logical approach to assess the value of shoeprint evidence. *Information Bulletin for Shoeprint/Toolmark Examiners* 8 (2): 15–25.

Taroni, F., Champod, C., and Margot, P. (1998). Forerunners of Bayesianism in early forensic science. *Jurimetrics Journal* 38: 183–200.

Taroni, F., De March, I., Garbolino, P., and Bozza, S. (2018a). Prova genetica del DNA e risultati dissonanti: come valutare congiuntamente gli elementi scientifici di prova. *Diritto Penale Contemporaneo* 11: 77–94.

Taroni, F., Garbolino, P., Biedermann, A., Aitken, C.G.G., and Bozza, S. (2018b). Reconciliation of subjective probabilities and frequencies in forensic science. *Law, Probability and Risk* 17: 243–264.

Taroni, F., Lambert, J.A., Fereday, L., and Werrett, D.J. (2002). Evaluation and presentation of forensic DNA evidence in European laboratories. *Science & Justice* 42: 21–28.

Taroni, F. and Margot, P. (2000). Fingerprint evidence evaluation: is it really so different to other evidence types? (Letter to the Editor). *Science & Justice* 40: 277–280.

Taroni, F. and Margot, P. (2001). General comments on the scale of conclusions in shoemarks – the need for a logical framework. *Information Bulletin for Shoeprint/Toolmark Examiners* 7 (2): 37–41.

Taylor, D. (2014). Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics* 11: 144–153.

Taylor, D., Abarno, D., Hicks, T.N., and Champod, C. (2016a). Evaluating forensic biology results given

source level propositions. *Forensic Science International: Genetics* 21: 54–67.

Taylor, D., Abarno, D., Rowe, E., and Rask-Nielsen, L. (2016b). Observations of DNA transfer within an operational forensic biology laboratory. *Forensic Science International: Genetics* 23: 33–49.

Taylor, D., Hicks, T.N., and Champod, C. (2016c). Using sensitivity analyses in Bayesian networks to highlight the impact of data paucity and direct future analyses: a contribution to the debate on measuring and reporting the precision of likelihood ratios. *Science & Justice* 56: 402–410.

Taylor, D., Biedermann, A., Hicks, T.N., and Champod, C. (2018a). A template for constructing Bayesian networks in forensic biology cases when considering activity level propositions. *Forensic Science International: Genetics* 33: 136–146.

Taylor, D., Curran, J.M., and Buckleton, J.S. (2018b). Likelihood ratio development for mixed Y-STR profiles. *Forensic Science International: Genetics* 35: 82–96.

Taylor, D., Kokshoorn, B., and Biedermann, A. (2018c). Evaluation of forensic genetics findings given activity level propositions: a review. *Forensic Science International: Genetics* 36: 34–49.

Taylor, D., Biedermann, A., Samie, L., Pun, K.M., Hicks, T.N., and Champod, C. (2017a). Helping to distinguish primary from secondary transfer events for trace DNA. *Forensic Science International: Genetics* 28: 155–177.

Taylor, D., Curran, J.M., and Buckleton, J.S. (2017b). Importance sampling allows H_d true tests of highly discriminating DNA profiles. *Forensic Science International: Genetics* 27: 74–81.

Taylor, D., Bright, J.A., Buckleton, J.S., and Curran, J.M. (2014). An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. *Forensic Science International: Genetics* 11: 56–63.

Taylor, D., Buckleton, J.S., and Evett, I.W. (2015). Testing likelihood ratios produced from complex DNA profiles. *Forensic Science International: Genetics* 16: 165–171.

Taylor, D., Samie, L., and Champod, C. (2019). Using Bayesian networks to track DNA movement through complex transfer scenarios. *Forensic Science International: Genetics* 42: 69–80.

Thagart, P. (2003). Why wasn't O.J. convicted? Emotional coherence and legal inference. *Cognition and Emotion* 17: 361–383.

Thompson, W.C. (1989). Are juries competent to evaluate statistical evidence? *Law and Contemporary Problems* 52: 9–41.

Thompson, W.C. (1995). Subjective interpretation, laboratory error and the value of DNA evidence: three case studies. *Genetica* 96: 153–168.

Thompson, W.C. (1997). Accepting lower standards: the National Research Council's second report on forensic DNA evidence. *Jurimetrics Journal* 37: 405–424.

Thompson, W.C. (2012). Discussion paper: Hard cases make bad law – reactions to *R v T*. *Law, Probability and Risk* 11: 347–359.

Thompson, W.C. and Ford, S. (1989). DNA typing: acceptance and weight of the new genetic identification tests. *Virginia Law Review* 75: 45–108.

Thompson, W.C., Grady, R.H., Lai, R., and Stern, H.S. (2018). Perceived strength of forensic scientists' reporting statements about source conclusions. *Law, Probability and Risk* 17: 133–155.

Thompson, W.C. and Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials. The prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behaviour* 11 167–187.

Thompson, W.C. and Scurich, N. (2018). When does absence of evidence constitute evidence of absence? *Forensic Science International* 291: e18–e19.

Thompson, W.C., Taroni, F., and Aitken, C.G.G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences* 48: 47–54.

Thompson, Y. and Williams, R. (1991). Blood group frequencies of the population of Trinidad and Tobago,

West Indies. *Journal of the Forensic Science Society* 31: 441–447.

Tippett, C.F., Emerson, V.J., Fereday, M.J., Lawton, F., and Lampert, S.M. (1968). The evidential value of the comparison of paint flakes from sources other than vehicles. *Journal of the Forensic Science Society* 8: 61–65.

Tribe, L. (1971). Trial by mathematics: precision and ritual in the legal process. *Harvard Law Review* 84: 1329–1393.

Triggs, C.M. and Buckleton, J.S. (2002). Logical implications of applying the principles of population genetics to the interpretation of DNA profiling evidence. *Forensic Science International* 128: 108–114.

Triggs, C.M. and Buckleton, J.S. (2003). The two-trace problem re-examined. *Science & Justice* 43: 127–134.

Triggs, C.M., Curran, J.M., Buckleton, J.S., and Walsh, K.A.J. (1997). The grouping problem in forensic glass analysis: a divisive approach. *Forensic Science International* 85: 1–14.

Triggs, C.M., Harbison, S.A., and Buckleton, J.S. (2000). The calculation of DNA match probabilities in mixed race populations. *Science & Justice* 40: 33–38.

Tully, G. and Stockdale, M. (2019). Commentary on: Hak (2019). Evaluation of the Forensic Science Regulator's recommendations regarding image comparison evidence. *Forensic Science International: Synergy* 1: 294–297; *Forensic Science International: Synergy* 1: 298–301.

Tvedebrink, T., Eriksen, P.S., Curran, J.M., Mogensen, H.S., and Morling, N. (2012). Analysis of matches and partial-matches in a Danish STR data set. *Forensic Science International: Genetics* 6: 387–392.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science* 185: 1124–1131.

Tversky, A. and Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In: *Judgement Under Uncertainty: Heuristics and Biases* (ed. D. Kahneman,

P. Slovic, and A. Tversky). 117–128 Cambridge: Cambridge University Press.

Tzidony, D. and Ravreboy, M. (1992). A statistical approach to drug sampling: a case study. *Journal of Forensic Sciences* 37: 1541–1549.

UN Office on Drugs and Crime (2009). *Guidelines on Representative Drug Sampling*. Vienna. https://www.unodc.org/documents/scientific/Drug_Sampling.pdf.

van den Hout, A. and Alberink, I. (2016). Posterior distributions for likelihood ratios in forensic science. *Science & Justice* 56: 397–401.

van Es, A., Wiarda, W., Hordijk, M., Alberink, I., and Vergeer, P. (2017). Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. *Science & Justice* 57: 181–192.

van Leeuwen, D.A. and Brummer, N. (2007). An introduction to application-independent evaluation of speaker recognition systems. In: *Speaker Classifications I* (ed. C. Muller), 330–353. Heidelberg: Springer-Verlag.

van Leeuwen, D.A. and Brümmer, N. (2013). The distribution of calibrated likelihood ratios in speaker recognition. https://arxiv.org/pdf/1304.1199.pdf.

van Oorschot, R.A.H., Szkuta, B., Ballantyne, K.N., and Goray, M. (2017). Need for dedicated training, competency assessment, authorisations and ongoing proficiency testing for those addressing DNA transfer issues. *Forensic Science International: Genetics* 6: e32–e34.

van Oorschot, R.A.H., Szkuta, B., Meakin, G.E., Kokshoorn, B., and Goray, M. (2019). DNA transfer in forensic science: a review. *Forensic Science International: Genetics* 38: 140–166.

Venables, W.M. and Ripley, B.D. (2002). *Modern Applied Statistics with S-Plus*, 4e. New York: Springer-Verlag.

Vicard, P. and Dawid, A.P. (2003). Estimating mutation rates from paternity data. In: *Atti del Convegno Modelli Complessi e metodi computazionali intensivi per la stima e*

la previsione, 415–418. Italy: Universitá Cá Foscari di Venezia.

Vicard, P., Dawid, A.P., Mortera, J., and Lauritzen, S.L. (2008). Estimating mutation rates from paternity casework. *Forensic Science International: Genetics* 2: 9–18.

Vincent, F.H.R. (2010). Inquiry into the circumstances that led to the conviction of Mr Farah Abdulkadir Jama. *Parliamentary paper, session 2006-2010, no. 301*, Victoria (Parliament), Melbourne.

Vito, G.F. and Latessa, E.J. (1989). *Statistical Applications in Criminal Justice*. London: Sage Publications.

Vuille, J., Biedermann, A., and Taroni, F. (2013). The importance of having a logical framework for expert conclusions in forensic DNA profiling. In: *Wrongful Convictions and Miscarriages of Justice* (ed. C.R. Huff and M. Killias), 137–159. New York: Routledge.

Wakefield, J.C., Skene, A.M., Smith, A.F.M., and Evett, I.W. (1991). The evaluation of fibre transfer evidence in forensic science: a case study in statistical modelling. *Applied Statistics* 40: 461–476.

Walsh, K.A.J. and Buckleton, J.S. (1986). On the problem of assessing the evidential value of glass fragments embedded in footwear. *Journal of the Forensic Science Society* 26: 55–60.

Walsh, K.A.J. and Buckleton, J.S. (1988). A discussion of the law of mutual independence and its application to blood group frequency. *Journal of the Forensic Science Society* 28: 95–98.

Walsh, S.J. and Buckleton, J.S. (2009). DNA databases and evidentiary issues. In: *Wiley Encyclopedia of Forensic Science* (ed. A. Jamieson and A. Moenssens), 831–839. Chichester: Wiley.

Walsh, K.A.J., Buckleton, J.S., and Triggs, C.M. (1996). A practical example of the interpretation of glass evidence. *Science & Justice* 36: 213–218.

Wasserstein, R.L. and Lazar, N.A. (2016). The ASA's statement on p-values: context, process and purpose. *The American Statistician* 70: 129–133.

Wasserstein, R.L., Schirm, A.L., and Lazar, N.A. (2019). Moving to a world beyond '$p < 0.05$'? *The American Statistician* 73 (S1): 1–19.

Watt, R., Roux, C., and Robertson, J. (2005). The population of coloured textile fibres in domestic washing machines. *Science & Justice* 45: 75–83.

Weir, B.S. (1992). Population genetics in the forensic DNA debate. *Proceedings of the National Academy of Sciences of the United States of America* 89: 11654–11659.

Weir, B.S. (1996). *Genetic Data Analysis II*. Sunderland: Sinauer Associates.

Weir, B.S. (1998). The coancestry coefficient. *Proceedings of the 8th International Symposium on Human Identification*. Madison, WI, USA: Promega Corporation, pp. 87–91.

Weir, B.S. (2000a). The consequences of defending DNA statistics. In: *Statistical Science in the Courtroom* (ed. J.L. Gastwirth), 86–97. New York: Springer-Verlag.

Weir, B.S. (2000b). Statistical analysis. In: *Encyclopedia of Forensic Science* (ed. J.A. Siegel, P.J. Saukko, and G.C. Knupfer), 545–550. San Diego, CA: Academic Press.

Weir, B.S. (2004). Matching and partially-matching DNA profiles. *Journal of Forensic Sciences* 49: 1009–1014.

Weir, B.S. (2007). Forensics. In: *Handbook of Statistical Genetics*, 3e (ed. D.J. Balding, M. Bishop, and C. Cannings), 1368–1392. Chichester: Wiley.

Weir, B.S. and Evett, I.W. (1992). Whose DNA? *American Journal of Human Genetics* 50: 869.

Weir, B.S. and Hill, W.G. (1993). Population genetics of DNA profiles. *Journal of the Forensic Science Society* 33: 218–225.

Weiss, C. (2003). Expressing scientific uncertainty. *Law, Probability and Risk* 2: 35–46.

Welch, B.L. (1937). The significance of the difference between two means when the population means are unequal. *Biometrika* 29: 350–362.

Weusten, J.J.A.M. (2011). Representative drug sampling: sample size calculations revisited. *Journal of Forensic Sciences* 56: 501–505.

Weyermann, C., Bucher, L., Majcherczyk, P., Mazzella, W., Roux, C., and Esseiva, P. (2012). Statistical discrimination of black gel pen inks analysed by laser desorption/ionization mass spectrometry. *Forensic Science International* 217: 127–133.

Wiggins, K.G., Drummond, P., and Hicks, T.N. (2004). A study in relation to the random distribution of four fibre types on clothing (incorporating a review of previous target fibre studies). *Science & Justice* 44: 141–148.

Wiggins, K.G., Emes, A., and Brackley, L.H. (2002). The transfer and persistence of small fragments of polyurethane foam onto clothing. *Science & Justice* 42: 105–110.

Wigmore, J.H. (1937). *The Science of Judicial Proof: As Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials*, 3e. Boston, MA: Little, Brown and Company.

Willuweit, S. and Roewer, L. (2007). Y chromosome haplotype reference database (YHRD): update. *Forensic Science International: Genetics* 1: 83–87.

Wilson, A., Aitken, C.G.G., Sleeman, R., and Carter, J. (2014). The evaluation of evidence relating to traces of cocaine on banknotes. *Forensic Science International* 236: 67–76.

Wilson, A., Aitken, C.G.G., Sleeman, R., and Carter, J. (2015). The evaluation of evidence for auto-correlated data in relation to traces of cocaine on banknotes. *Applied Statistics* 64: 275–298.

Winkler, R.L. (1996). *An Introduction to Bayesian Inference and Decision*. Gainesville, FL: Probabilistic Publishing.

Wixted, J.T., Christenfeld, N.J.S., and Rouder, J.N. (2019). Calculating the posterior odds from a single-match DNA database search. *Law, Probability and Risk* 18: 1–23.

Wixted, J.T. and Rouder, J.N. (2019). Calculating the posterior odds from a single-match DNA database search with hidden assumptions. *Law, Probability and Risk* 18: 229–234.

Wooley, J.R. (1991). A response to Lander: the courtroom perspective. *American Journal of Human Genetics* 49: 892–893.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* 56: 330–338.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* 15: 323–354.

Wright, S. (1965). The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution* 19: 395–420.

Zabell, S. (1976). Book review – probability and statistical inference in ancient and medieval Jewish literature – by N.L. Rabinovitch (1973), University of Toronto Press, Toronto, Canada. *Journal of the American Statistical Association* 71: 996–998.

Zadora, G., Martyna, A., Ramos-Castro, D., and Aitken, C.G.G. (2014). *Statistical Analysis in Forensic Science. Evidential Value of Multivariate Physicochemical Data*. Chichester: Wiley.

Zamengo, L., Frison, G., Gregio, M., Orrù, G., and Sciarrone, R. (2011). Determination of illicit drugs in seized material: Role of sampling and analysis in estimation of measurement uncertainty. *Forensic Science International* 208: 108–123. https://doi.org/10.1016/j.forsciint.2010.11.018.

Zeisel, H. and Kaye, D.H. (1997). *Prove it with Figures*. New York: Springer-Verlag.

Zynda, L. (2016). Subjectivism. In: *The Oxford Handbook of Probability and Philosophy* (ed. A. Hajek and C. Hitchcock), 360–381. Oxford: Oxford University Press.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

Notation

The Greek and Roman alphabets provide a large choice of letters to be used for mathematical notation. Despite this large choice, some letters, such as x, are used in this book to mean more than one thing. It is hoped that no letter or symbol is asked to mean more than one thing at the same time and that the list that follows will help readers to know what each letter or symbol does mean at any particular point. Chapter or Section references are given for the first or main use of many of the letters or symbols.

$\in$:	denotes containment in a set, thus $x \in \{1, 2, 3\}$ indicates x is one of the integers 1, 2, or 3.
$\ldots$:	three dots, written on the line, indicate 'and so on in sequence to'. Thus $x_1, \ldots, x_5$ can be read as 'x_1 and so on in sequence to x_5' and is short-hand for the sequence x_1, x_2, x_3, x_4, x_5. Usually, the last subscript is a

general one such as n so that a sequence of n items would be written as $x_1, \ldots, x_n$.

$\cdots$: three dots, written above the line, indicate 'a repeat of the operation immediately before and after the dots'. Thus $x_1 + \cdots + x_5$ is shorthand for '$x_1 + x_2 + x_3 + x_4 + x_5$'. Similarly $x_1 \times \cdots \times x_5$ is shorthand for '$x_1 \times x_2 \times x_3 \times x_4 \times x_5$'. Usually, the last subscript is a general one such as n so that a sum or product of n items would be written as $x_1 + \cdots + x_n$ or $x_1 \times \cdots \times x_n$. Also, the symbol $\times$ is often omitted and $x_1 \times \cdots \times x_n$ written as $x_1 \cdots x_n$ or

$$\Pr(R \mid S) \cdot \Pr(S) = \Pr(R \mid S) \times \Pr(S).$$

$(\ldots)$: For numerical ranges, the limits are not included in the range. Thus $\theta \in (0, 1)$ is equivalent to $0 < \theta < 1$. Contrast this with $[\ldots]$; Section 4.7.2.

$[\ldots]$: For numerical ranges, the limits are included in the range. Thus $\theta \in [0, 1]$ is equivalent to $0 \leq \theta \leq 1$. Contrast this with $(\ldots)$; Section 4.7.2.

$\sum$: the sum of terms following the symbol. For example, $\sum_{i=1}^{n} x_i$ denotes the sum of $x_1, \ldots, x_n$ $(x_1 + \cdots + x_n)$, Section A.2.2.

$\cdot$: a single dot written as a suffix denotes summation over the index indicated by the location of the suffix; for example, the mean over j of $(x_{ij}, i = 1, \ldots, m, j = 1, \ldots, n)$ could be indicated $\bar{x}_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$; Section 7.6.5.

$\prod$: the product of terms following the symbol. For example, $\prod_{i=1}^{n} Pr(S_i)$ denotes the product of the probabilities $Pr(S_1), \ldots, Pr(S_n)$ $(Pr(S_1) \cdots Pr(S_n))$, Section 3.5.4.

$\bar{}$: to be read as 'the opposite of' or the 'complement of', thus if M denotes male, $\bar{M}$ denotes female; Section 2.1.1.

$\bar{}$: to be read as 'the mean of', thus $\bar{x}$ is the mean of a set of measurements $x_1, \ldots, x_n$; Section A.3.1.

$\equiv$: to be read as 'as equivalent to'. For example, if M denotes male and F denotes female, then $M \equiv \bar{F}$ and $F \equiv \bar{M}$.

$>>$: to be read as 'is very much greater than' (in contrast to $>$ which is simply 'is greater than'); Section 3.6.2.

$\propto$: to be read as 'is proportional to'. For example, this is often used in Bayesian analysis where the distribution of the random

	variable is taken to be proportional to an expression involving only terms in the random variable and omitting other terms which are needed to ensure the distribution is a probability distribution, i.e. has a total probability of 1. Use of such a notation eases the algebraic manipulations associated with Bayesian inference; Section A.3.3.
$(\mathbf{X} \mid \boldsymbol{\theta}, \Sigma) \sim N(\boldsymbol{\theta}, \Sigma)$:	multivariate random variable $\mathbf{X}$ has a Normal distribution with mean vector $\boldsymbol{\theta}$ and covariance matrix Σ; Section A.3.9.
$\mid x \mid$:	for x a number, the absolute value of x; if $x > 0$, $\mid x \mid = x$, ; if $x < 0$, $\mid x \mid = -x$; for example, $\mid 6 \mid = 6$, $\mid -6 \mid = 6$.
$\mid \Sigma \mid$:	determinant of the matrix Σ.
$\sim$:	is distributed as; $X \sim N(\mu, \sigma^2)$ is X is distributed Normally with mean μ and variance σ^2.
$\sim$:	for consequences, indicates indifference, thus $c_1 \sim c_2$ indicates indifference between consequences c_1 and c_2; Section 2.8.1.
$\prec$:	for consequences, indicates a strict preference, thus $c_1 \prec c_2$

	indicates a strict preference for c_2 over c_1; Section 2.8.1.
$\preceq$:	for consequences, a lack of preference, thus $c_1 \preceq c_2$ indicates that c_1 is not preferred to c_2; Section 2.8.1.
α:	a prior parameter for the beta distribution; Section A.3.7.
$1-\alpha$:	the size of a Bayesian credible interval; Section 4.1.
α_1:	given evidence E the posterior probability of proposition H_1, $\Pr(H_1 \mid E)$; Section 2.8.2.
α_2:	given evidence E the posterior probability of proposition H_2, $\Pr(H_2 \mid E)$; Section 2.8.2.
b_0:	probability of zero groups of material being found as background; Section 5.3.2.1.
$b_{g,\mathbf{m}}$:	probability of g groups of material of sizes $(m_1, \ldots, m_g) = \mathbf{m}$ being found; Section 5.3.2.1.
$b^*_{1,m}$:	probability of occurrence of one group of material from some external source other than that of the investigation; Section 5.3.2.5.
β:	a prior parameter for the beta distribution; Section A.3.7.
$B(\alpha, \beta)$:	the normalising constant for a beta distribution; Section A.3.7.

$B(\alpha_1, \ldots, \alpha_k)$:	the normalising constant for a Dirichlet distribution; Section A.3.8.
$Be(\alpha, \beta)$:	beta distribution with parameters α and β; Section A.3.7.
c_2:	$\Pr(\bar{E}^* \mid \bar{E})$; Section 5.6.2.1.
$c(d, \theta)$:	the consequence of decision d when the actual state of nature is θ, sometimes abbreviated to c; Section 2.8.1.
$\mathcal{C}$:	the set of all consequences $c(d, \theta)$; Section 2.8.1.
$C_X(Y)$:	the probability a person of blood group Y innocently bears a bloodstain of blood group X; Section 6.2.4.
d:	decision; Section 2.8.1.
d:	distance; Section 7.8.
$\mathcal{D}$:	set of all decisions; Section 2.8.1.
Γ:	analytical result from the inspection of trace evidence.
γ:	frequency of Γ in a relevant population; a parameter.
$\Gamma(x + 1)$:	the gamma function; Section A.3.5.
E:	quality or measurements of evidential material; Chapter 1.
E_c:	quality or measurements of evidential material of control form, also denoted x; Section 1.7.1.

E_r:	quality or measurements of evidential material of recovered form, also denoted y; Section 1.7.1.
Ev:	the totality of the evidence, equals (M, E); Section 1.7.1.
E^*:	a report about E; Section 5.6.2.1.
EL:	expected loss; Section 2.8.1.
EU:	expected utility; Section 2.8.1.
$E(\theta_i)$:	the mean of the variable θ_i, also known as the *expectation*; Section A.3.8.
f_1:	$\Pr(E^* \mid \bar{E})$; Section 5.6.2.1.
$f_{t,z-1}\{.\}$:	the probability density function of the $t-$ distribution with $(z-1)$ degrees of freedom; Section A.3.4.
g:	number of groups; Section 5.3.2.1.
h_1:	$\Pr(E^* \mid E)$; Section 5.6.2.1.
H_d:	the proposition of the defence; Section 2.1.1.
H_p:	the proposition of the prosecution; Section 2.1.1.
I:	the identity matrix; Section A.1.6.
I:	background information; Section 1.7.9.
$I_{s,t}$:	information available to person S at time t: Section 1.7.5.
l_i:	$i = 1, 2$, loss function; Section 2.8.2.

$\mathrm{L}\,(c(d, \theta))$:	the loss for a given consequence $c(d, \theta)$ of decision d with true state of nature θ; Section 2.8.1.
$\log_e$:	logarithm to base e.
$\log_{10}$:	logarithm to base 10.
m_2:	$\Pr(\bar{E}^* \mid E)$; Section 5.6.2.1.
M:	evidential material; Section 1.7.1.
M_c:	evidential material of control form; Section 1.7.1.
M_r:	evidential material of recovered form; Section 1.7.1.
m:	number of items inspected; Section 4.3.
μ:	mean or expectation of a Normal distribution; Section A.3.2.
n:	number of items not inspected; Section 4.3.
n:	number of groups transferred between two objects; Section 5.3.2.1.
N:	consignment size ($= m + n$); Section 4.3.
$N(\theta, \sigma^2)$:	Normal distribution of mean θ and standard deviation σ.
ν:	degrees of freedom; Section A.3.4.
O:	odds; Section 2.1.
Ω:	the universal set, $\Pr(\Omega) = 1.$; Section 1.7.6.
Ω:	population of characteristics of items; Section 7.6.2.

P: the probability of what is observed, or anything more extreme, calculated assuming the null hypothesis is true, known as the *significance probability*: Section 3.6.1.

P_i: $i = 1, 2$, population; Section 7.7.

p: probability of transfer of material to the person of interest from the crime scene or from the person of interest to the crime scene, persisting and being recovered if the person of interest were innocent; Section 5.3.3.3.

p: the number of variables in a vector, e.g. $\mathbf{x}^T = (x_1, \ldots x_p)$.

ϕ: the empty set, $\Pr(\phi) = 0$; Section 1.7.6

p_i: probability of presence of i (≥ 0) groups of material on the person of interest; Section 6.2.2.

Pr: probability.

Ψ: subpopulations; Section 2.2.2.

π: $\{\pi_1, \ldots, \pi_n\}$ is a permutation of $\{1, \ldots, n\}$; Section 1.7.7.

π_1: the prior probability that the person of interest is the source of a recovered stain; Section 6.1.4.1.

π_i: ($i = 2, \ldots, n$), the prior probabilities of the alternative propositions for other members of

	a relevant population other than a person of interest; Section 6.1.4.1.
Q:	random variable corresponding to quantity to be estimated; Section 4.6.
q:	equals $1 - p$ where p is the relative frequency in a sample.
q:	quantity to be estimated; Section 4.6.
ρ:	population correlation coefficient; Section A.3.9.
R:	number of items in the consignment which are illicit; Section 4.3.2.
r:	the probability of relevance: some, all or none of the transferred material may be present for innocent reasons (e.g. reasons unassociated with the offender) and some, none or all for guilty reasons (e.g. reasons associated with the offender); some of the material is selected for analysis. If the selected material is part of that which was there for guilty reasons then it is defined as relevant; Section 5.3.3.2.
s:	standard deviation of a sample or of measured items; Section A.3.1.

s_l: probability that a group of fragments found on members of a population is large; Section 6.2.2.

σ: standard deviation of a population; Section A.3.1.

Σ: covariance matrix; Section A.3.9.

$t_\nu(P)$: the $100P\%$ point of the t-distribution with ν degrees of freedom; Section A.3.4.

t_n: probability of transfer of $n(\geq 0)$ items of material to the person of interest from the crime scene or from the person of interest to the crime scene, persisting and being recovered if H_p is true. Section 6.2.2.

t_W: the numerator of the Student's t-density (7.10); Section 7.3.1.

t'_n: probability of transfer of $n\ (\geq 0)$ items of material to the offender from the crime scene or from the offender to the crime scene, persisting and being recovered if H_d is true; Section 5.3.2.1.

θ: probability of at least one match of evidence of a given frequency with an identified individual in a population of individuals unrelated to the identified individual and of finite size; Section 2.5.5.

θ: a state of nature, Section 2.8.1.

θ:	parameter of a probability distribution; for a prior distribution it is treated as a variable; Sections A.3.7 and A.3.8.
θ:	proportion of the consignment which contains illicit items; Section 4.3.1.
θ:	co-ancestry coefficient F_{ST}; Section 6.1.5.
$\tilde{\theta}$:	point estimate of parameter θ; Section 4.7.
θ^*:	optimal decision; Section 4.7.
θ_0:	lower bound for the proportion of the consignment which contains illicit items; Section 4.3.1.
Θ:	the set of all possible states of nature; Section 2.8.1.
Θ:	set of all possible values of θ; Section 4.7.
$U(c(d, \theta))$:	the utility of the consequence of decision d for true state of nature θ; Section 2.8.1.
$u_{n\|T}$:	transfer probability of n pieces of evidence, conditional on event T; Section 5.3.2.5.
V:	the value of evidence, the likelihood ratio; Section 2.4.1.
$V_s(x)$:	the value of the evidence, the likelihood ratio for x versus s; Section 6.1.6.3.

w_j: the weight of the contents of the j-th item not examined which is illicit; Section 4.6.

$\bar{w}$: mean weight of items not inspected which are illicit; Section 4.6.

x: measurement on control material; Section 7.3.

x_i: the weight of the contents of the i-th item examined which is illicit; Section 4.6.

$\bar{x}$: mean weight of inspected items which are illicit; Section 4.6.

$\mathbf{x}_{ij}$: background multivariate data for sample j in group i, $i = 1, \ldots, m$, $j = 1, \ldots, n$; Section 7.6.2.

$\mathbf{x}_{ijk}$: background multivariate data for replicate j of member k in group i, $i = 1, \ldots, m$, $j = 1, \ldots, n$, $k = 1, \ldots, t$; Section 7.6.5.

$x!$: x factorial; when x is a positive integer, the product of x with all positive integers less than it and greater than zero, $= x(x-1)(x-2)\cdots 2 \cdot 1$; conventionally $0! = 1$; Section A.2.1.

y: measurement on recovered material; Section 7.3.

y: number of items not inspected which are illicit; this number is unknown and modelled by a

	beta-binomial distribution; Section 4.3.
$\mathbf{y}_1$:	multivariate control data; Section 7.6.
$\mathbf{y}_2$:	multivariate recovered data; Section 7.6.
z_i:	a member of the background data for univariate data, $i = 1, \ldots, k$; Section 7.5.
z:	number of items inspected which are found to be illicit; $z \leq m$; Section 4.3.
z:	realisation of a random variable Z which has a standard normal distribution; $Z \sim N(0, 1)$.
$\binom{n}{x}$:	the binomial coefficient, the number of ways in which x $(0 \leq x \leq n)$ items may be chosen from n $(\geq x)$ in which no attention is paid to ordering; equals $n!/\{x!(n-x)!\}$; Section A.2.1.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

Cases

Amanda Knox (2015) Corte Suprema di Cassazione – sentenza n. 1105 UP 27.3.2015, dep. 7.9.2015, R.G.N. 32598/2014; Section 5.2.

Commonwealth of Massachusetts v. Nicola Sacco and Bartolomeo Vanzetti, 1921; Sections 1.2 and 2.9.

Commonwealth v. McClellan, 178 A.3d 874 (Pa. Super. Ct. 2018); Section 2.4.2.

Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579, 1993; Section 8.2.

E. Ross v. State of Indiana, Indiana Court of Appeal, 13 May, 1996; Sections 2.5.2, 2.5.6 and 2.5.11.

Ex parte De La Cruz 466 S.W.3d 855 (Tex. Crim. App. 2015); Section 2.5.10.4.

Ghebrezghi v. State, No. 2432 (Md. Ct. Spec. App. Mar. 15, 2018); Section 2.4.2.

Holmes v. South Carolina 547 U.S. 319 (2006) 126 S.Ct 1727; Section 2.5.10.4.

In re Michelle I., 189 A.D. 2d., 998, 1000 (N.Y. App. Div. 1993; Section 2.5.10.4.

Robocast, Inc. v. Microsoft Corp. Civil Action No. 10-1055-RGA (D. Del. Jan. 13, 2014); Section 2.5.9.

Ross v. State, B14-90-00659, Tex. App. 13 Feb.,1992; Sections 2.5.6 and 2.5.11.

State v. Davis, 814 S.W. 2d 593, Mo., 1991; Section 2.5.2.

State v. Glover, 825 S.W. 2d 127, Tex. Crim. App., 1992; Section 2.5.2.

State v. Hollywood, 680 P. 2d 655, 657, Or. Ct. App., 1984; Section 2.5.10.4.

State v. Klindt, District Court of Scott County, Iowa, Case number 115422, 1968; Section 6.1.4.2.

State of New Jersey vs. J.M. Spann, 130 N.J. 484; 617 A. 2d 247, 1993; Section 6.3.4.

State v. Wright, 253 P.3d 838 (Mont. 2011); Section 2.5.8.

Swearingen v. State 101 S.W.3d 89 (Tex. Crim. App. 2003); Section 2.5.10.4.

U.S. v. Davis Case No.: 4:18-cr-00011 (W.D. Va. Sep. 11, 2019); Sections 2.5.9, 2.5.11.

U.S. v. Gissantner, United States District Court Western District of Michigan Southern Division, Case n. 1:17-cr-130, decided Oct 16, 2019; Section 6.1.5.

U.S. v. Jakobetz, 955 F. 2d 786 2nd Cir. 1992; Section 2.5.11.

U.S. v. Morrow, United States District Court, D. Columbia, 374 F. Supp. 2d 51 (D.D.C. 2005); Section 2.5.10.3.

U.S. v. Pirre, 927, F. 2d 694, 2nd Cir., 1991; Section 4.6.

U.S. v. Shonubi: Shonubi V: 962 F.Supp.370 (E.D.N.Y.1997); *Shonubi IV*: 103 F.3d 1085 (2d Cir. 1997); *Shonubi III*: 895 F.Supp. 460 (E.D.N.Y. 1995); *Shonubi II*: 998 F.2d 84 (2d Cir. 1993); *Shonubi I*: 802 F.Supp. 859 (E.D.N.Y. 1992); Section 4.3.

U.S. v. Tibbs, Superior Court of the District of Columbia, Criminal Division Case No. 2016 CF1 19431 (2019); Section 2.5.9.

U.S. v. Williams Case No. 3:13-cr-00764-WHO-1 (N.D. Cal. Apr. 29, 2019); Section 2.4.2.

Vermont v. Streich, T., 658 A.2d 38, 1995; Section 2.5.2.

Wike v. State, 596 So 2nd 1020, Fla S. Ct., 1992; Section 2.5.2.

Wilson v. Maryland, Court of Appeals of Maryland, 370 Md 191, 803 A. 2d 1034, August 4th, 2002; Sections 2.5.1 and 3.4.

Yara Gambirasio (2016) Corte di Assise di Bergamo – sentenza di primo grado n. 1/16 Sent. Assise, dep. 27.9.2016, n. 7701/14 R.G.N.R.; Section 2.5.6.

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

Author Index

Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

Subject Index