Bayesian Approaches to Clinical Trials and Health-Care Evaluation

David J. Spiegelhalter
MRC Biostatistics Unit, Cambridge, UK

Keith R. Abrams
University of Leicester, UK

Jonathan P. Myles
Cancer Research UK, London, UK

John Wiley & Sons, Ltd

Bayesian Approaches to Clinical Trials and Health-Care Evaluation

STATISTICS IN PRACTICE

Advisory Editor

Stephen Senn
University of Glasgow, 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.

Bayesian Approaches to Clinical Trials and Health-Care Evaluation

David J. Spiegelhalter
MRC Biostatistics Unit, Cambridge, UK

Keith R. Abrams
University of Leicester, UK

Jonathan P. Myles
Cancer Research UK, London, UK

John Wiley & Sons, Ltd

The work is based on an original NHS Health Technology Assessment funded project (93/50/05). Adapted with kind permission of the National Coordinating Centre for Health Technology Assessment.

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-49975-7

Typeset in 10/12 pt Photina from LaTeX files supplied by the author, processed by Kolam Information Services Pvt. Ltd, Pondicherry, India.
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

This book began life as a review of Bayesian methods in health technology assessment commissioned by the UK National Health Service Research and Development Programme, which appeared as Spiegelhalter *et al.* (2000). It was then thought to be a good idea to change the review into a basic introduction to Bayesian methods which also tried to cover the field of clinical trials and health-care evaluation. We did not realise the amount of work this would involve.

We are very grateful to all those who have read all or part of the manuscript and given such generous comments, particularly David Jones, Laurence Freedman, Mahesh Parmar, Tony Ades, Julian Higgins, Nicola Cooper, Cosetta Minelli, Alex Sutton and Denise Kendrick. Unfortunately, by tradition, we must take full responsibility for all errors and idiosyncrasies. Our particular thanks go to Daniel Farewell for writing the BANDY program, and Nick Freemantle for providing data. The University of Leicester provided the second author with study leave, during which part of this work was carried out. Finally, we must thank Rob Calver and Siân Jones at Wiley for being so patient with the repeated excuses for delay: in the words of Douglas Adams (1952–2001). ''I love deadlines. I especially like the whooshing sound they make as they go flying by''. We hope it has been worth the wait.

List of Examples

1

Introduction

1.1 WHAT ARE BAYESIAN METHODS?

Bayesian statistics began with a posthumous publication in 1763 by Thomas Bayes, a Nonconformist minister from the small English town of Tunbridge Wells. His work was formalised as *Bayes theorem* which, when expressed mathematically, is a simple and uncontroversial result in probability theory. However, specific uses of the theorem have been the subject of continued controversy for over a century, giving rise to a steady stream of polemical arguments in a number of disciplines. In recent years a more balanced and pragmatic perspective has developed and this more ecumenical attitude is reflected in the approach taken in this book: we emphasise the benefits of Bayesian analysis and spend little time criticising more traditional statistical methods.

The basic idea of Bayesian analysis is reasonably straightforward. Suppose an unknown quantity of interest is the median years of survival gained by using an innovative rather than a standard therapy on a defined group of patients: we shall call this the 'treatment effect'. A clinical trial is carried out, following which conventional statistical analysis of the results would typically produce a *P*-value for the null hypothesis that the treatment effect is zero, as well as a point estimate and a confidence interval as summaries of what this particular trial tells us about the treatment effect. A Bayesian analysis supplements this by focusing on how the trial should change our opinion about the treatment effect. This perspective forces the analyst to explicitly state

- a reasonable opinion concerning the plausibility of different values of the treatment effect *excluding* the evidence from the trial (known as the prior distribution),
- the support for different values of the treatment effect based *solely* on data from the trial (known as the likelihood),

and to combine these two sources to produce

- a final opinion about the treatment effect (known as the posterior distribution).

Bayesian Approaches to Clinical Trials and Health-Care Evaluation D. J. Spiegelhalter, K. R. Abrams and J. P. Myles
 ISBN: 0-471-49975-7

The final combination is done using Bayes theorem, which essentially weights the likelihood from the trial with the relative plausibilities defined by the prior distribution. This basic idea forms the entire foundation of Bayesian analysis, and will be developed in stages throughout the book.

One can view the Bayesian approach as a formalisation of the process of learning from experience, which is a fundamental characteristic of all scientific investigation. Advances in health-care typically happen through incremental gains in knowledge rather than paradigm-shifting breakthroughs, and so this domain appears particularly amenable to a Bayesian perspective.

1.2 WHAT DO WE MEAN BY 'HEALTH-CARE EVALUATION'?

Our concern is with the evaluation of 'health-care interventions', which is a deliberately generic term chosen to encompass all methods used to improve health, whether drugs, medical devices, health education programmes, alternative systems for delivering care, and so on. The appropriate evaluation of such interventions is clearly of deep concern to individual consumers, health-care professionals, organisations delivering care, policy-makers and regulators: such evaluations are commonly called 'health-technology assessments', but we feel this term carries connotations of 'high' technology that we wish to avoid.

A wide variety of research designs have been used in evaluation, and it is not the purpose of this book to argue the benefits of one design over another. Rather, we are concerned with appropriate methods for analysing and interpreting evidence from one or multiple studies of possibly varying designs. Many of the standard methods of analysis revolve around the classical randomised controlled trial (RCT): these include power calculations at the design stage, methods for controlling Type I error within sequential monitoring, calculation of *P*-values and confidence intervals at the final analysis, and meta-analytic techniques for pooling the results of multiple studies. Such methods have served the medical research community well.

The increasing sophistication of evaluations is, however, highlighting the limitations of these traditional methods. For example, when carrying out a clinical trial, the many sources of evidence and judgement available beforehand may be inadequately summarised by a single 'alternative hypothesis', monitoring may be complicated by simultaneous publication of related studies, and multiple subgroups may need to be analysed and reported. Randomised trials may not be feasible or may take a long time to reach conclusions. A single clinical trial will also rarely be sufficient to inform a policy decision, such as embarking or continuing on a research programme, regulatory approval of a drug or device, or recommendation of a treatment at an individual or population level. Standard statistical methods are designed for summarising the evidence from single studies or pooling evidence from similar studies, and have difficulties dealing with the pervading complexity of multiple sources of evidence. Many have argued that a fresh, Bayesian, approach is worth investigating.

1.3 A BAYESIAN APPROACH TO EVALUATION

We may define a Bayesian approach as 'the explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a health-care evaluation'. The argument of this book is that such a perspective can be more *flexible* than traditional methods in that it can adapt to each unique situation, more *efficient* in using all available evidence, more *useful* in providing predictions and inputs for making decisions for specific patients, for planning research or for public policy, and more *ethical* in both clarifying the basis for randomisation and fully exploiting the experience provided by past patients.

For example, a Bayesian approach allows evidence from diverse sources to be pooled through assuming that their underlying probability models (their likelihoods) share parameters of interest: thus the 'true' underlying effect of an intervention may feature in models for both randomised trials and observational data, even though there may be additional adjustments for potential biases, different populations, crossovers between treatments, and so on.

Attitudes have changed since Feinstein (1977) claimed that 'a statistical consultant who proposes a Bayesian analysis should therefore be expected to obtain a suitably informed consent from the clinical client whose data are to be subjected to the experiment'. Increasing attention to the Bayesian approach is shown by the medical and statistical literature, the popular scientific press, pharmaceutical companies and regulatory agencies. However, many important outstanding questions remain: in particular, to what extent will the scientific community, or the regulatory authorities, allow the explicit introduction of evidence that is not totally derived from observed data, or the formal pooling of data from studies of differing designs? Indeed, Berry (2001) warns that 'There is as much Bayesian junk as there is frequentist junk. Actually, there's probably more of the former because, to the uninitiated, the Bayesian approach seems like it provides a free lunch'. External evidence must therefore be introduced with caution, and used in a clear, explicit and transparent manner that can be challenged by those who need to critique any analysis: this balanced approach should help resolve these complex questions.

1.4 THE AIM OF THIS BOOK AND THE INTENDED AUDIENCE

This book is intended to provide:

- a review of the essential ideas of Bayesian analysis as applied to the evaluation of health-care interventions, without obscuring the essential message with undue technicalities;
- a suggested 'template' for reporting a Bayesian analysis;

- a critical commentary on similarities and differences between Bayesian and conventional approaches;
- a structured review of published work in the areas covered;
- a wide range of stand-alone examples of Bayesian methods applied to real data, mainly in a common format, with accompanying software which will allow the reader to reproduce all analyses;
- a guide to potential areas where Bayesian methods might be particularly valuable, and where further research may be necessary;
- an indication of appropriate methods that may be applied in different contexts (although this is not intended as a 'cookbook');
- a range of exercises suitable for use in a course based on the material in this book.

Our intended audience comprises anyone with a good grasp of quantitative methods in health-care evaluation, and whose mathematical and statistical training includes basic calculus and probability theory, use of normal tables, clinical trial design, and familiarity with hypothesis testing, estimation, confidence intervals, and interpretation of odds and hazard ratios, up to the level necessary to use standard statistical packages. Bayesian statistics has a (largely deserved) reputation for being mathematically challenging and difficult to put into practice, although we recommend O'Hagan and Luce (2003) as a good non-technical preliminary introduction to the basic ideas. In this book we deliberately try to use the simplest possible analytic methods, largely based on normal distributions, without distorting the conclusions: more technical aspects are placed in starred sections that can be omitted without loss of continuity. There is a steady progression throughout the book in terms of analytic complexity, so that by the final chapters we are dealing with methods that are at the research frontier. We hope that readers will find their own level of comfort and make some effort to transcend it.

1.5 STRUCTURE OF THE BOOK

We have struggled to decide on an appropriate structure for the material in this book. It could be ordered by *stage of evaluation* and so separate, for example, initial observational studies, RCTs possibly for licensing purposes, cost-effectiveness analysis and monitoring interventions in routine use. Alternatively, we might structure by *study design*, with discussion of randomised trials, databases, case–control studies, and so on. Finally, we could identify the *modelling issue*, for example prior distributions, alternative forms for likelihoods, and loss functions. We have, after much deliberation, made a compromise and used aspects of all three proposals, using extensive examples to weave together analytic techniques with evaluation problems.

Chapter 2 is a brief *revision* of important aspects of traditional statistical analysis, covering issues such as probability distributions, normal tables, parameterisation of outcomes, summarising results by estimates and confidence intervals, hypothesis testing and sample-size assessment. There is a particular emphasis on normal likelihoods, since they are an important prerequisite for much of the subsequent Bayesian analysis, but we also provide a fairly detailed catalogue of other distributions and their use.

Chapter 3 forms the core of the book, being an *overview* of the main features of the Bayesian approach. Topics include the subjective interpretation of probability, use of prior to posterior analysis in a clinical trial, assessing the evidence in reported clinical trial results, comparing hypotheses, predictions, decision-making, exchangeability and hierarchical models, and computation: these topics are then applied to substantive problems in later chapters. Differing perspectives on prior distributions and loss functions are shown to lead to different schools of Bayesianism. A proposed checklist for reporting Bayesian health-care evaluations forms the basis for all further examples in the book.

Chapter 4 briefly critiques the 'classical' statistical approach to health-care evaluation and makes a *comparison* with the Bayesian approach. Hypothesis tests, *P*-values, Bayes factors, stopping rules and the 'likelihood principle' are discussed with examples. This chapter can be skipped without loss of continuity.

Chapter 5 deals in detail with sources of *prior distributions*, such as expert opinion, summaries of evidence, 'off-the-shelf' default priors and hierarchical priors based on exchangeability assumptions. The criticism of prior opinions in the light of data is featured, and a detailed taxonomy provided of ways of using historical data as a basis for prior opinion.

Chapter 6 attempts to structure the substantial work on Bayesian approaches to all aspects of RCTs, including design, monitoring, reporting, and interpretation. The many worked examples emphasise the need for analysis of sensitivity to alternative prior assumptions.

Chapter 7 covers *observational studies*, such as case–control and other non-randomised designs. Particular aspects emphasised include the explicit modelling of potential biases with such designs, and non-randomised comparisons of institutions including ranking into 'league tables'.

Chapter 8 considers the *synthesis of evidence* from multiple studies, starting from 'standard' meta-analysis and then considering various extensions such as potential dependence of treatment effects on baseline risk. We particularly focus on examples of 'generalised evidence synthesis', which might feature studies of different designs, or 'indirect' comparison of treatments that have never been directly compared in a trial.

Chapter 9 examines how Bayesian analyses may be used to inform *policy*, including cost-effectiveness analysis, research planning and regulatory affairs. The view of alternative stakeholders is emphasised, as is the integration of evidence synthesis and cost-effectiveness in a single unified analytic model.

Chapter 10 includes a final summary, general discussion and some suggestions for future research. Appendix A briefly describes available software and Internet sites of interest.

Most of the chapters finish with a list of key points and questions/exercises, and some have a further guide to the literature.

This structure will inevitably mean some overlap in methodological questions, such as the appropriate form of the prior distribution, and whether it is reasonable to adopt an explicit loss function. For example, a particular issue that arises in many contexts is the appropriate means of including historical data. This will be introduced as a general issue and a list of different approaches provided (Section 3.16), and then these approaches will be illustrated in four different contexts in which one might wish to use historical data: first, obtaining a prior distribution from historical studies (Section 5.4); second, historical controls in randomised trials (Section 6.9); third, modelling the potential biases in observational studies (Section 7.3), and fourth, pooling data from many sources in an evidence synthesis (Section 8.2). This overlap means that a considerable amount of cross-referencing is inevitable and ideally there would be hypertext links, but a traditional book format forces us into a linear structure.

Different audiences may want to focus on different parts of the book. The material up to Chapter 5 comprises a basic short course in Bayesian analysis, suitable for both students and researchers. After that, Chapter 6 may be of more interest to statisticians working with clinical trials in the pharmaceutical industry or the public sector, while Chapters 7–9 may be more appropriate for those exploring policy decisions. However, there are no clear boundaries and we hope that most of the material is relevant for much of the potential readership.

In order to avoid disappointment, we should make clear what this book does *not* contain:

- There is almost no guidance on data analysis, model checking and many other essential ingredients of professional statistical practice. Our discussion of study design is limited to sample-size calculations, and there is little contribution to the debate concerning the relative importance of observational and randomised studies.
- There is no rigorous mathematical or philosophical development of the Bayesian approach, and the technical development is limited entirely to the level required for the examples.
- The examples are almost all taken from published work by ourselves and others, and although they deal with real problems and use real data, there is necessarily a degree of simplification in the presentation. In addition, while the Bayesian approach emphasises the formal use of substantive knowledge and subjective opinion, it is inevitable that judgements are introduced in a somewhat stylised manner into such 'second-hand' examples. We should also point out that numbers given in the text have been rounded, and the accompanying programs should be used for a more accurate analysis.

- There is limited development of the decision-theoretic approach to evaluation, and many will feel this is a serious omission. This bias arises from two related reasons. First, our personal experience has been almost entirely concerned with problems of inference, and so that is what we feel qualified to write about. Second, it will become clear that we have some misgivings concerning the application of decision theory in this context, and so prefer to emphasise the more immediately relevant material.
- There is very limited exploration of more general Bayesian approaches to modelling data that arise in health-care evaluations, such as applications to survival analysis, longitudinal models, non-compliance in trials, drop-outs and other missing data, and so on.

The accompanying website will be found at `http://www.mrc-bsu.cam.ac.uk/bayeseval/`, which provides code for most of the examples in the book, either using the BANDY spreadsheet program for simple analysis of odds and hazard ratios, or WinBUGS code for more complex examples. The website will also contain a list of any errors detected.

Finally, we should emphasise that this book is not intended as a polemic in favour of Bayesianism – there have been enough of those – and we shall try to avoid making exaggerated claims as to the benefits of this new 'treatment' for statistical problems. Our hope is that we can contribute to the responsible use of Bayesian methods and hence help in a small way towards the development of cost-effective health-care.

2

Basic Concepts from Traditional Statistical Analysis

The Bayesian approach, to a considerable extent, supplements rather than replaces the kind of analyses traditionally carried out in assessing health-care interventions, and in this chapter we shall briefly review some of the basic ideas that will subsequently be found useful. In particular, probability theory is fundamental to Bayesian analysis, and we therefore revise the basic concepts with a natural emphasis on Bayes theorem. We also consider random variables and probability distributions with particular emphasis on the normal distribution, which plays a vital role in summarising what the observed data can tell us about unknown quantities of interest. A particularly important practical aspect is the transformation of output from standard statistical packages into a form amenable to Bayesian interpretation.

Bayesian analysis makes a much wider use of probability distributions than traditional statistical methods, in that not only are sampling distributions required for summaries of data, but also a wide range of distributions are used to represent prior opinion about proportions, event rates, and other unknown quantities. The *shapes* of distributions therefore become particularly important, as they are intended to represent the plausibility of different values, and so we shall provide (in starred sections) extensive graphical displays as well the usual formulae.

Most of the issues addressed in this chapter are covered in a concise and readable manner in standard textbooks such as Altman (2001) and Berry *et al.* (2001b). In addition, Clayton and Hills (1993) consider a likelihood-based approach to many of the models that are frequently encountered in epidemiology and health-care evaluation.

Bayesian Approaches to Clinical Trials and Health-Care Evaluation D. J. Spiegelhalter, K. R. Abrams and J. P. Myles
 ISBN: 0-471-49975-7

2.1 PROBABILITY

2.1.1 What is probability?

Suppose a is some event which may or may not take place, such as the next toss of a coin coming up heads. Although we may casually speak of the 'probability' of a occurring, and give it a mathematical notation $p(a)$, it is perhaps remarkable that there is no universally agreed definition of what this term means. Perhaps the currently most accepted interpretation is the following: $p(a)$ is the proportion of times a will occur in an infinitely long series of repeated identical situations. This is known as the 'frequentist' perspective, as it rests on the frequency with which specific events occur. However, a number of other interpretations of probability have been made throughout history, and we shall consider a different, 'subjective', definition in Section 3.1.

There is little dispute, however, about the mathematical properties of probability. Let a and b be events, and H represent the context in which a and b might arise, and let $p(a|H)$ denote the probability of a given the context H: the vertical line represents 'conditioning'. Then $p(a|H)$ is a number that satisfies the following three basic rules:

1. *Bounds*.

$$0 \leq p(a|H) \leq 1,$$

   where $p(a|H) = 0$ if a is impossible and $p(a|H) = 1$ if a is certain in the context H.

2. *Addition rule*. If a and b are mutually exclusive (*i.e.* one at most can occur),

$$p(a \text{ or } b|H) = p(a|H) + p(b|H).$$

   (We note that, for technical reasons, it is helpful if Rule 2 is taken as holding for an infinite set of mutually exclusive events.)

3. *Multiplication rule*. For any events a and b,

$$p(a \text{ and } b|H) = p(a|b,H)p(b|H).$$

We say that a and b are independent if $p(a \text{ and } b|H) = p(a|H)p(b|H)$ or equivalently $p(a|b,H) = p(a|H)$: thus the fact that b has occurred does not alter the probability of a. The multiplication rule can equivalently be expressed as the definition of conditional probability,

$$p(a|b,H) = \frac{p(a \text{ and } b|H)}{p(b|H)},$$

provided $p(b|H) \neq 0$.

The explicit introduction of the context H is unusual in standard texts and we shall subsequently drop it to avoid accusations of pedantry: however, it is always useful to keep in mind that *all probabilities are conditional* and so, if the situation changes, then probabilities may change. We shall see in Section 3.1 that this notion forms the basis of *subjective probability*, in which H, the context, represents the information on which an individual bases their *own* subjective assessment of the *degree of belief, i.e.* probability, of an event occurring.

Example 2.1 illustrates that these rules can be given an immediate intuitive justification by comparison with a standard experiment.

Example 2.1 *Dice: Illustration of rules of probability*

Suppose H denotes the roll of two perfectly balanced six-sided dice, and let '$\equiv$' denote 'is equivalent to'.

Rule 1. For a single die: if $a \equiv$ 'throw 7', then $p(a) = 0$; if $a \equiv$ 'throw ≤ 6', then $p(a) = 1$. If c is the sum of the two dice: then if $c \equiv$ '13', then $p(c) = 0$; if $c \equiv$ '≤ 12', then $p(c) = 1$.

Rule 2. For a single die: if $a \equiv$ 'throw 3', $b \equiv$ 'throw 4', then

$$\begin{aligned} p(a \text{ or } b) &= p(a) + p(b) \text{ since } a \text{ and } b \text{ are mutually exclusive} \\ &= 1/6 + 1/6 = 1/3. \end{aligned}$$

Rule 3. If we throw two dice: if $a \equiv$ 'first die throw 2', $b \equiv$ 'second die throw 5', then

$$\begin{aligned} p(a \text{ and } b) &= p(a)p(b) \text{ since } a \text{ and } b \text{ are independent} \\ &= 1/6 \times 1/6 = 1/36. \end{aligned}$$

If $a \equiv$ 'total score of the two throws is greater than or equal to 6', $b \equiv$ 'first die throw 1', then

$$\begin{aligned} p(a \text{ and } b) &= p(a|b)p(b) \\ &= 1/3 \times 1/6 = 1/18. \end{aligned}$$

Suppose we also consider the events 'a and b' and 'a and $\overline{b}$', where $\overline{b}$ represents the event 'not b'. Then 'a and b' and 'a and $\overline{b}$' are mutually exclusive and together form the event a, and hence, using Rule 2, we have the identity

$$p(a) = p(a \text{ and } b) + p(a \text{ and } \overline{b}) \tag{2.1}$$

which is known as 'marginalisation'. Further, by using Rule 3, we obtain

$$p(a) = p(a|b)p(b) + p(a|\overline{b})p(\overline{b}), \tag{2.2}$$

which is known by the curious title of 'extending the conversation' (or 'extending the argument'). Example 2.2 shows these expressions follow naturally from considering the full 'joint' distribution over all possible combinations of events.

Example 2.2 *Prognosis: Marginalisation and extending the conversation*

Suppose we wish to determine the probability of survival (up to a specified point in time) following a particular cancer diagnosis, given that it depends on the stage of disease at diagnosis amongst other factors. Whilst directly specifying the probability of surviving, denoted b, may be difficult, by extending the conversation to include whether the cancer was at an early stage, denoted a, or not, denoted $\overline{a}$, we obtain from (2.1),

$$p(b) = p(b|a)p(a) + p(b|\overline{a})p(\overline{a}).$$

For example, suppose patients with early stage disease have a good prognosis, say $p(b|a) = 0.80$, but for late stage it is poor, say $p(b|\overline{a}) = 0.20$, and that of new diagnoses the majority, 90%, are early stage, *i.e.* $p(a) = 0.90$ and $p(\overline{a}) = 0.10$. Then the marginal probability of surviving is $p(b) = 0.80 \times 0.90 + 0.20 \times 0.10 = 0.74$.

Table 2.1 shows all possible combinations of events and their probabilities, as well as the marginal probabilities that, appropriately, appear in the margin of the table. The joint probabilities of events have been obtained by Rule 2 so that, for example, $p(b \text{ and } a) = p(b|a)p(a) = 0.80 \times 0.90 = 0.72$.

Table 2.1 Probabilities of all combinations of survival and stage, including marginal probabilities.

	Early stage a	Late stage $\overline{a}$	
Survive b	0.72	0.02	0.74
Not survive $\overline{b}$	0.18	0.08	0.26
	0.90	0.10	1.00

2.1.2 Odds and log-odds

Any probability p can also be expressed in terms of 'odds' O, where

$$O = \frac{p}{1-p}$$

and

$$p = \frac{O}{1+O},$$

so that, for example, a probability of 0.20 (20% chance) corresponds to odds of $O = 0.20/0.80 = 0.25$ or, in betting parlance, '4 to 1 against'. Conversely, betting odds of '7 to 4 against' correspond to $O = 4/7$, or a probability of $p = 4/11 = 0.36$.

The natural logarithm (denoted log) of the odds is termed the 'logit', so that

$$\text{logit}(p) = \log\left[\frac{p}{1-p}\right].$$

2.1.3 Bayes theorem for simple events

A number of properties can immediately be derived from Rules 1 to 3 of Section 2.1.1. Since $p(b \text{ and } a) = p(a \text{ and } b)$, Rule 3 implies that $p(b|a)p(a) = p(a|b)p(b)$, or equivalently

$$p(b|a) = \frac{p(a|b)}{p(a)} \times p(b). \tag{2.3}$$

We have proved Bayes theorem! In words, this vital result tells us how an initial probability $p(b)$ is changed into a conditional probability $p(b|a)$ when taking into account the event a occurring: it should be clear by this description that we are interpreting Bayes theorem as providing a formal mechanism for learning from experience.

Equation (2.3) also holds for $\overline{b}$, so that

$$p(\overline{b}|a) = \frac{p(a|\overline{b})}{p(a)} \times p(\overline{b}), \tag{2.4}$$

and dividing (2.3) by (2.4) we obtain the *odds form* for Bayes theorem:

$$\frac{p(b|a)}{p(\overline{b}|a)} = \frac{p(a|b)}{p(a|\overline{b})} \times \frac{p(b)}{p(\overline{b})}. \tag{2.5}$$

Thus $p(b)/p(\overline{b}) = p(b)/(1 - p(b))$, the odds on b before taking into account the event a, which is changed into the new odds $p(b|a)/p(\overline{b}|a)$ after conditioning on a. Equation (2.5) shows how Bayes theorem accomplishes this transformation without even explicitly calculating $p(a)$, and this insight is exploited in Section 3.2.

Example 2.3 *Prognosis (continued): Bayes theorem for single events*

Suppose we were given Table 2.1, and wanted to use Bayes theorem to tell us how knowing the stage of the disease at diagnosis revises our probability for survival a. Initially, before we know the stage, $p(b) = 0.74$ from the

marginal probability in Table 2.1. Suppose we find out that the disease is at an early stage, *i.e.* a, where we know from Table 2.1 that $p(a|b) = 0.72/0.74 = 0.97$ and $p(a) = 0.9$. Hence from (2.3) we obtain a revised probability of survival

$$p(b|a) = \frac{0.97}{0.9} \times 0.74 = 0.80,$$

matching what, in fact, we knew already.

To use the odds form of Bayes theorem (2.5) we first require the initial odds for survival, *i.e.* $p(b)/p(\overline{b}) = 0.74/0.26 = 2.85$, and the ratio $p(a|b)/p(a|\overline{b}) = 0.97/0.69 = 1.405$. Then from (2.5) we obtain the final odds on survival as $2.85 \times 1.41 = 4.01$, corresponding to a probability $p(b|a) = 0.80$ (up to rounding error).

The two forms of Bayes theorem both give the required results and can be thought of as a means of moving from a marginal probability in a table to a conditional probability having taken into account some evidence. As we shall see in Section 3.2, it is this use of Bayes theorem that is used in many diagnostic testing situations without any controversy.

2.2 RANDOM VARIABLES, PARAMETERS AND LIKELIHOOD

2.2.1 Random variables and their distributions

Random variables have a somewhat complex formal definition, but it is sufficient to think of them as unknown quantities that may take on one of a set of values: traditionally a random variable is denoted by a capital Latin letter, say Y, before being observed and by a lower-case letter y as a specific observed value. This convention tends to be broken in Bayesian analysis, in which all unknown quantities are considered as random variables, but we shall try to keep to it where it clarifies the exposition.

Loosely speaking, $p(y)$ denotes the probability of a random variable Y taking on each of its possible values y. $p(y)$ is formally known as the *probability density function*, and the probability that Y does not exceed y, $P(Y \leqslant y)$, is termed the *probability distribution function*. We shall tend to use 'probability distribution' as a generic term, hopefully without causing confusion.

Probability distributions may be:

Binary. When Y can take on one of two values, we shall generally use the notation $Y = 1$ for when an event of interest occurs, and $Y = 0$ when it does not: this is