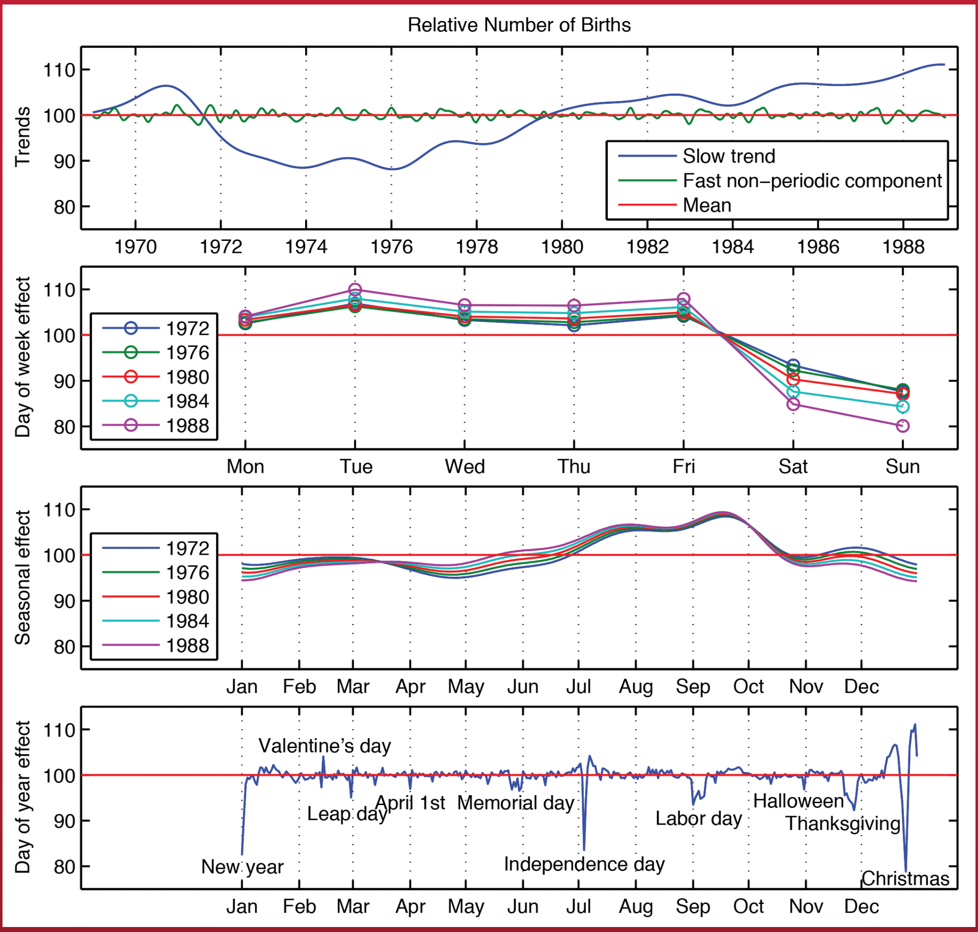
Texts in Statistical Science

Bayesian Data Analysis

Third Edition



Andrew Gelman, John B. Carlin, Hal S. Stern,
David B. Dunson, Aki Vehtari, and Donald B. Rubin

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Bayesian Data Analysis

Third Edition

CHAPMAN & HALL/CRC
Texts in Statistical Science Series

Analysis of Failure and Survival Data
P. J. Smith

The Analysis of Time Series: An Introduction, Sixth Edition
C. Chatfield

Applied Bayesian Forecasting and Time Series Analysis
A. Pole, M. West, and J. Harrison

Applied Categorical and Count Data Analysis
W. Tang, H. He, and X.M. Tu

Applied Nonparametric Statistical Methods, Fourth Edition
P. Sprent and N.C. Smeeton

Applied Statistics: Handbook of GENSTAT Analyses
E.J. Snell and H. Simpson

Applied Statistics: Principles and Examples
D.R. Cox and E.J. Snell

Applied Stochastic Modelling, Second Edition
B.J.T. Morgan

Bayesian Data Analysis, Third Edition
A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin

Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians
R. Christensen, W. Johnson, A. Branscum, and T.E. Hanson

Bayesian Methods for Data Analysis, Third Edition
B.P. Carlin and T.A. Louis

Beyond ANOVA: Basics of Applied Statistics
R.G. Miller, Jr.

The BUGS Book: A Practical Introduction to Bayesian Analysis
D. Lunn, C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter

A Course in Categorical Data Analysis
T. Leonard

A Course in Large Sample Theory
T.S. Ferguson

Data Driven Statistical Methods
P. Sprent

Decision Analysis: A Bayesian Approach
J.Q. Smith

Design and Analysis of Experiments with SAS
J. Lawson

Elementary Applications of Probability Theory, Second Edition
H.C. Tuckwell

Elements of Simulation
B.J.T. Morgan

Epidemiology: Study Design and Data Analysis, Third Edition
M. Woodward

Essential Statistics, Fourth Edition
D.A.G. Rees

Exercises and Solutions in Statistical Theory
L.L. Kupper, B.H. Neelon, and S.M. O'Brien

Exercises and Solutions in Biostatistical Theory
L.L. Kupper, B.H. Neelon, and S.M. O'Brien

Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models
J.J. Faraway

A First Course in Linear Model Theory
N. Ravishanker and D.K. Dey

Generalized Additive Models: An Introduction with R
S. Wood

Generalized Linear Mixed Models: Modern Concepts, Methods and Applications
W. W. Stroup

Graphics for Statistics and Data Analysis with R
K.J. Keen

Interpreting Data: A First Course in Statistics
A.J.B. Anderson

Introduction to General and Generalized Linear Models
H. Madsen and P. Thyregod

An Introduction to Generalized Linear Models, Third Edition
A.J. Dobson and A.G. Barnett

Introduction to Multivariate Analysis
C. Chatfield and A.J. Collins

Introduction to Optimization Methods and Their Applications in Statistics
B.S. Everitt

Introduction to Probability with R
K. Baclawski

Introduction to Randomized Controlled Clinical Trials, Second Edition
J.N.S. Matthews

Introduction to Statistical Inference and Its Applications with R
M.W. Trosset

Introduction to Statistical Limit Theory
A.M. Polansky

Introduction to Statistical Methods for Clinical Trials
T.D. Cook and D.L. DeMets

Introduction to Statistical Process Control
P. Qiu

Introduction to the Theory of Statistical Inference
H. Liero and S. Zwanzig

Large Sample Methods in Statistics
P.K. Sen and J. da Motta Singer

Linear Algebra and Matrix Analysis for Statistics
S. Banerjee and A. Roy

Logistic Regression Models
J.M. Hilbe

Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition
D. Gamerman and H.F. Lopes

Mathematical Statistics
K. Knight

Modeling and Analysis of Stochastic Systems, Second Edition
V.G. Kulkarni

Modelling Binary Data, Second Edition
D. Collett

Modelling Survival Data in Medical Research, Second Edition
D. Collett

Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioural Scientists
D.J. Hand and C.C. Taylor

Multivariate Statistics: A Practical Approach
B. Flury and H. Riedwyl

Multivariate Survival Analysis and Competing Risks
M. Crowder

Nonparametric Methods in Statistics with SAS Applications
O. Korosteleva

Pólya Urn Models
H. Mahmoud

Practical Data Analysis for Designed Experiments
B.S. Yandell

Practical Longitudinal Data Analysis
D.J. Hand and M. Crowder

Practical Multivariate Analysis, Fifth Edition
A. Afifi, S. May, and V.A. Clark

Practical Statistics for Medical Research
D.G. Altman

A Primer on Linear Models
J.F. Monahan

Principles of Uncertainty
J.B. Kadane

Probability: Methods and Measurement
A. O'Hagan

Problem Solving: A Statistician's Guide, Second Edition
C. Chatfield

Randomization, Bootstrap and Monte Carlo Methods in Biology, Third Edition
B.F.J. Manly

Readings in Decision Analysis
S. French

Sampling Methodologies with Applications
P.S.R.S. Rao

Stationary Stochastic Processes: Theory and Applications
G. Lindgren

Statistical Analysis of Reliability Data
M.J. Crowder, A.C. Kimber, T.J. Sweeting, and R.L. Smith

Statistical Methods for Spatial Data Analysis
O. Schabenberger and C.A. Gotway

Statistical Methods for SPC and TQM
D. Bissell

Statistical Methods in Agriculture and Experimental Biology, Second Edition
R. Mead, R.N. Curnow, and A.M. Hasted

Statistical Process Control: Theory and Practice, Third Edition
G.B. Wetherill and D.W. Brown

Statistical Theory: A Concise Introduction
F. Abramovich and Y. Ritov

Statistical Theory, Fourth Edition
B.W. Lindgren

Statistics for Accountants
S. Letchford

Statistics for Epidemiology
N.P. Jewell

Statistics for Technology: A Course in Applied Statistics, Third Edition
C. Chatfield

Statistics in Engineering: A Practical Approach
A.V. Metcalfe

Statistics in Research and Development, Second Edition
R. Caulcutt

Stochastic Processes: An Introduction, Second Edition
P.W. Jones and P. Smith

Survival Analysis Using S: Analysis of Time-to-Event Data
M. Tableman and J.S. Kim

The Theory of Linear Models
B. Jørgensen

Time Series Analysis
H. Madsen

Time Series: Modeling, Computation, and Inference
R. Prado and M. West

Understanding Advanced Statistical Methods
P.H. Westfall and K.S.S. Henning

Texts in Statistical Science

Bayesian Data Analysis

Third Edition

Andrew Gelman
John B. Carlin
Hal S. Stern
David B. Dunson
Aki Vehtari
Donald B. Rubin

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150506

International Standard Book Number-13: 978-1-4398-9820-8 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Bayesian Data Analysis
Third edition
(with corrections for third printing, 7 Apr 2015)

Andrew Gelman
Columbia University

John B. Carlin
University of Melbourne

Hal S. Stern
University of California, Irvine

David B. Dunson
Duke University

Aki Vehtari
Aalto University

Donald B. Rubin
Harvard University

This page intentionally left blank

Contents

Preface

This book is intended to have three roles and to serve three associated audiences: an introductory text on Bayesian inference starting from first principles, a graduate text on effective current approaches to Bayesian modeling and computation in statistics and related fields, and a handbook of Bayesian methods in applied statistics for general users of and researchers in applied statistics. Although introductory in its early sections, the book is definitely not elementary in the sense of a first text in statistics. The mathematics used in our book is basic probability and statistics, elementary calculus, and linear algebra. A review of probability notation is given in Chapter 1 along with a more detailed list of topics assumed to have been studied. The practical orientation of the book means that the reader's previous experience in probability, statistics, and linear algebra should ideally have included strong computational components.

To write an introductory text alone would leave many readers with only a taste of the conceptual elements but no guidance for venturing into genuine practical applications, beyond those where Bayesian methods agree essentially with standard non-Bayesian analyses. On the other hand, we feel it would be a mistake to present the advanced methods without first introducing the basic concepts from our data-analytic perspective. Furthermore, due to the nature of applied statistics, a text on current Bayesian methodology would be incomplete without a variety of worked examples drawn from real applications. To avoid cluttering the main narrative, *there are bibliographic notes at the end of each chapter* and references at the end of the book.

Examples of real statistical analyses appear throughout the book, and we hope thereby to give an applied flavor to the entire development. Indeed, given the conceptual simplicity of the Bayesian approach, it is only in the intricacy of specific applications that novelty arises. Non-Bayesian approaches dominated statistical theory and practice for most of the last century, but the last few decades have seen a re-emergence of Bayesian methods. This has been driven more by the availability of new computational techniques than by what many would see as the theoretical and logical advantages of Bayesian thinking.

In our treatment of Bayesian inference, we focus on practice rather than philosophy. We demonstrate our attitudes via examples that have arisen in the applied research of ourselves and others. Chapter 1 presents our views on the foundations of probability as empirical and measurable; see in particular Sections 1.4–1.7.

Changes for the third edition

The biggest change for this new edition is the addition of Chapters 20–23 on nonparametric modeling. Other major changes include weakly informative priors in Chapters 2, 5, and elsewhere; boundary-avoiding priors in Chapter 13; an updated discussion of cross-validation and predictive information criteria in the new Chapter 7; improved convergence monitoring and effective sample size calculations for iterative simulation in Chapter 11; presentations of Hamiltonian Monte Carlo, variational Bayes, and expectation propagation in Chapters 12 and 13; and new and revised code in Appendix C. We have made other changes throughout.

During the eighteen years since completing the first edition of *Bayesian Data Analysis*, we have worked on dozens of interesting applications which, for reasons of space, we are not able to add to this new edition. Many of these examples appear in our book, *Data Analysis*

Using Regression and Hierarchical/Multilevel Models, as well as in our published research articles.

We have made some small corrections and updates for the second printing of the third edition.

Online information

Additional materials, including the data used in the examples, solutions to many of the end-of-chapter exercises, and any errors found after the book goes to press, are posted at `http://www.stat.columbia.edu/~gelman/book/`. Feel free to send any comments to us directly.

Acknowledgments

We thank many students, colleagues, and friends for comments and advice and also acknowledge the public funding that made much of this work possible.

In particular, we thank Stephen Ansolabehere, Adriano Azevedo, Jarrett Barber, Richard Barker, Tom Belin, Michael Betancourt, Suzette Blanchard, Rob Calver, Brad Carlin, Bob Carpenter, Alicia Carriquiry, Samantha Cook, Alex Damour, Victor De Oliveira, Vince Dorie, David Draper, John Emerson, Steve Fienberg, Alex Franks, Byron Gajewski, Yuanjun Gao, Daniel Gianola, Yuri Goegebeur, David Hammill, Chad Heilig, Matt Hoffman, Chuanpu Hu, Zaiying Huang, Shane Jensen, Yoon-Sook Jeon, Pasi Jylanki, Jay Kadane, Jouni Kerman, Gary King, Lucien Le Cam, Yew Jin Lim, Rod Little, Tom Little, Chuanhai Liu, Xuecheng Liu, Tomoki Matsumoto, Peter McCullagh, Mary Sara McPeek, Xiao-Li Meng, Baback Moghaddam, Olivier Nimeskern, Peter Norvig, Ali Rahimi, Thomas Richardson, Christian Robert, Scott Schmidler, Matt Schofield, Andrea Siegel, Sandip Sinharay, Elizabeth Stuart, Dwight Sunada, Andrew Swift, Eric Tassone, Francis Tuerlinckx, Iven Van Mechelen, Amos Waterland, Rob Weiss, Lo-Hua Yuan, and Alan Zaslavsky. We especially thank John Boscardin, Jessica Hwang, Daniel Lee, Phillip Price, and Radford Neal.

This work was partially supported by research grants from the National Science Foundation, National Institutes of Health, Institute of Education Sciences, National Security Agency, Department of Energy, and Academy of Finland.

Many of our examples have appeared in books and articles written by ourselves and others, as we indicate in the bibliographic notes and exercises in the chapters where they appear.[1]

Finally, we thank Caroline, Nancy, Hara, Amy, Ilona, and other family and friends for their love and support during the writing and revision of this book.

[1]In particular: Figures 1.3–1.5 are adapted from the *Journal of the American Statistical Association* 90 (1995), pp. 696, 702, and 703, and are reprinted with permission of the American Statistical Association. Figures 2.6 and 2.7 come from Gelman, A., and Nolan, D., *Teaching Statistics: A Bag of Tricks*, Oxford University Press (1992), pp. 14 and 15, and are reprinted with permission of Oxford University Press. Figures 19.8–19.10 come from the *Journal of the American Statistical Association* 91 (1996), pp. 1407 and 1409, and are reprinted with permission of the American Statistical Association. Table 19.1 comes from Berry, D., *Statistics: A Bayesian Perspective*, first edition, copyright 1996 Wadsworth, a part of Cengage Learning, Inc. Reproduced by permission. www.cengage.com/permissions. Figures 18.1 and 18.2 come from the *Journal of the American Statistical Association* 93 (1998), pp. 851 and 853, and are reprinted with permission of the American Statistical Association. Figures 9.1–9.3 are adapted from the *Journal of Business and Economic Statistics* 21 (2003), pp. 219 and 223, and are reprinted with permission of the American Statistical Association. We thank Jack Taylor for the data used to produce Figure 23.4.

Part I: Fundamentals of Bayesian Inference

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations. In Chapters 1–3, we introduce several useful families of models and illustrate their application in the analysis of relatively simple data structures. Some mathematics arises in the analytical manipulation of the probability distributions, notably in transformation and integration in multiparameter problems. We differ somewhat from other introductions to Bayesian inference by emphasizing stochastic simulation, and the combination of mathematical analysis and simulation, as general methods for summarizing distributions. Chapter 4 outlines the fundamental connections between Bayesian and other approaches to statistical inference. The early chapters focus on simple examples to develop the basic ideas of Bayesian inference; examples in which the Bayesian approach makes a practical difference relative to more traditional approaches begin to appear in Chapter 3. The major practical advantages of the Bayesian approach appear in Chapter 5, where we introduce *hierarchical models*, which allow the parameters of a prior, or population, distribution themselves to be estimated from data.

This page intentionally left blank

Chapter 1

Probability and inference

1.1 The three steps of Bayesian data analysis

This book is concerned with practical methods for making inferences from data using probability models for quantities we observe and for quantities about which we wish to learn. The essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis.

The process of Bayesian data analysis can be idealized by dividing it into the following three steps:

1. Setting up a *full probability model*—a joint probability distribution for all observable and unobservable quantities in a problem. The model should be consistent with knowledge about the underlying scientific problem and the data collection process.
2. Conditioning on observed data: calculating and interpreting the appropriate *posterior distribution*—the conditional probability distribution of the unobserved quantities of ultimate interest, given the observed data.
3. Evaluating the fit of the model and the implications of the resulting posterior distribution: how well does the model fit the data, are the substantive conclusions reasonable, and how sensitive are the results to the modeling assumptions in step 1? In response, one can alter or expand the model and repeat the three steps.

Great advances in all these areas have been made in the last forty years, and many of these are reviewed and used in examples throughout the book. Our treatment covers all three steps, the second involving computational methodology and the third a delicate balance of technique and judgment, guided by the applied context of the problem. The first step remains a major stumbling block for much Bayesian analysis: just where do our models come from? How do we go about constructing appropriate probability specifications? We provide some guidance on these issues and illustrate the importance of the third step in retrospectively evaluating the fit of models. Along with the improved techniques available for computing conditional probability distributions in the second step, advances in carrying out the third step alleviate to some degree the need to assume correct model specification at the first attempt. In particular, the much-feared dependence of conclusions on 'subjective' prior distributions can be examined and explored.

A primary motivation for Bayesian thinking is that it facilitates a common-sense interpretation of statistical conclusions. For instance, a Bayesian (probability) interval for an unknown quantity of interest can be directly regarded as having a high probability of containing the unknown quantity, in contrast to a frequentist (confidence) interval, which may strictly be interpreted only in relation to a sequence of similar inferences that might be made in repeated practice. Recently in applied statistics, increased emphasis has been placed on interval estimation rather than hypothesis testing, and this provides a strong impetus to the Bayesian viewpoint, since it seems likely that most users of standard confidence intervals give them a common-sense Bayesian interpretation. One of our aims in this book is to indicate the extent to which Bayesian interpretations of common simple statistical procedures are justified.

Rather than argue the foundations of statistics—see the bibliographic note at the end of this chapter for references to foundational debates—we prefer to concentrate on the pragmatic advantages of the Bayesian framework, whose flexibility and generality allow it to cope with complex problems. The central feature of Bayesian inference, the direct quantification of uncertainty, means that there is no impediment in principle to fitting models with many parameters and complicated multilayered probability specifications. In practice, the problems are ones of setting up and computing with such large models, and a large part of this book focuses on recently developed and still developing techniques for handling these modeling and computational challenges. The freedom to set up complex models arises in large part from the fact that the Bayesian paradigm provides a conceptually simple method for coping with multiple parameters, as we discuss in detail from Chapter 3 on.

1.2 General notation for statistical inference

Statistical inference is concerned with drawing conclusions, from numerical data, about quantities that are not observed. For example, a clinical trial of a new cancer drug might be designed to compare the five-year survival probability in a population given the new drug to that in a population under standard treatment. These survival probabilities refer to a large *population* of patients, and it is neither feasible nor ethically acceptable to experiment on an entire population. Therefore inferences about the true probabilities and, in particular, their differences must be based on a *sample* of patients. In this example, even if it were possible to expose the entire population to one or the other treatment, it is never possible to expose anyone to both treatments, and therefore statistical inference would still be needed to assess the *causal inference*—the comparison between the observed outcome in each patient and that patient's unobserved outcome if exposed to the other treatment.

We distinguish between two kinds of *estimands*—unobserved quantities for which statistical inferences are made—first, potentially observable quantities, such as future observations of a process, or the outcome under the treatment not received in the clinical trial example; and second, quantities that are not directly observable, that is, parameters that govern the hypothetical process leading to the observed data (for example, regression coefficients). The distinction between these two kinds of estimands is not always precise, but is generally useful as a way of understanding how a statistical model for a particular problem fits into the real world.

Parameters, data, and predictions

As general notation, we let θ denote unobservable vector quantities or population *parameters* of interest (such as the probabilities of survival under each treatment for randomly chosen members of the population in the example of the clinical trial), y denote the observed data (such as the numbers of survivors and deaths in each treatment group), and $\tilde{y}$ denote unknown, but potentially observable, quantities (such as the outcomes of the patients under the other treatment, or the outcome under each of the treatments for a new patient similar to those already in the trial). In general these symbols represent multivariate quantities. We generally use Greek letters for parameters, lower case Roman letters for observed or observable scalars and vectors (and sometimes matrices), and upper case Roman letters for observed or observable matrices. When using matrix notation, we consider vectors as column vectors throughout; for example, if u is a vector with n components, then $u^T u$ is a scalar and uu^T an $n \times n$ matrix.

Observational units and variables

In many statistical studies, data are gathered on each of a set of n objects or *units*, and we can write the data as a vector, $y = (y_1, \dots, y_n)$. In the clinical trial example, we might label y_i as 1 if patient i is alive after five years or 0 if the patient dies. If several variables are measured on each unit, then each y_i is actually a vector, and the entire dataset y is a matrix (usually taken to have n rows). The y variables are called the 'outcomes' and are considered 'random' in the sense that, when making inferences, we wish to allow for the possibility that the observed values of the variables could have turned out otherwise, due to the sampling process and the natural variation of the population.

Exchangeability

The usual starting point of a statistical analysis is the (often tacit) assumption that the n values y_i may be regarded as *exchangeable*, meaning that we express uncertainty as a joint probability density $p(y_1, \dots, y_n)$ that is invariant to permutations of the indexes. A nonexchangeable model would be appropriate if information relevant to the outcome were conveyed in the unit indexes rather than by explanatory variables (see below). The idea of exchangeability is fundamental to statistics, and we return to it repeatedly throughout the book.

We commonly model data from an exchangeable distribution as independently and identically distributed (*iid*) given some unknown parameter vector θ with distribution $p(\theta)$. In the clinical trial example, we might model the outcomes y_i as iid, given θ, the unknown probability of survival.

Explanatory variables

It is common to have observations on each unit that we do not bother to model as random. In the clinical trial example, such variables might include the age and previous health status of each patient in the study. We call this second class of variables *explanatory variables*, or *covariates*, and label them x. We use X to denote the entire set of explanatory variables for all n units; if there are k explanatory variables, then X is a matrix with n rows and k columns. Treating X as random, the notion of exchangeability can be extended to require the distribution of the n values of $(x, y)_i$ to be unchanged by arbitrary permutations of the indexes. It is *always* appropriate to assume an exchangeable model after incorporating sufficient relevant information in X that the indexes can be thought of as randomly assigned. It follows from the assumption of exchangeability that the distribution of y, given x, is the same for all units in the study in the sense that if two units have the same value of x, then their distributions of y are the same. Any of the explanatory variables x can be moved into the y category if we wish to model them. We discuss the role of explanatory variables (also called predictors) in detail in Chapter 8 in the context of analyzing surveys, experiments, and observational studies, and in the later parts of this book in the context of regression models.

Hierarchical modeling

In Chapter 5 and subsequent chapters, we focus on *hierarchical models* (also called *multilevel models*), which are used when information is available on several different levels of observational units. In a hierarchical model, it is possible to speak of exchangeability at each level of units. For example, suppose two medical treatments are applied, in separate randomized experiments, to patients in several different cities. Then, if no other information were available, it would be reasonable to treat the patients within each city as exchangeable and also treat the results from different cities as themselves exchangeable. In practice it would make sense to include, as explanatory variables at the city level, whatever relevant information we have on each city, as well as the explanatory variables mentioned before at the individual level, and then the conditional distributions given these explanatory variables would be exchangeable.

1.3 Bayesian inference

Bayesian statistical conclusions about a parameter θ, or unobserved data $\tilde{y}$, are made in terms of *probability* statements. These probability statements are conditional on the observed value of y, and in our notation are written simply as $p(\theta|y)$ or $p(\tilde{y}|y)$. We also implicitly condition on the known values of any covariates, x. It is at the fundamental level of conditioning on observed data that Bayesian inference departs from the approach to statistical inference described in many textbooks, which is based on a retrospective evaluation of the procedure used to estimate θ (or $\tilde{y}$) over the distribution of possible y values conditional on the true unknown value of θ. Despite this difference, it will be seen that in many simple analyses, superficially similar conclusions result from the two approaches to statistical inference. However, analyses obtained using Bayesian methods can be easily extended to more complex problems. In this section, we present the basic mathematics and notation of Bayesian inference, followed in the next section by an example from genetics.

Probability notation

Some comments on notation are needed at this point. First, $p(\cdot|\cdot)$ denotes a conditional probability density with the arguments determined by the context, and similarly for $p(\cdot)$, which denotes a marginal distribution. We use the terms 'distribution' and 'density' interchangeably. The same notation is used for continuous density functions and discrete probability mass functions. Different distributions in the same equation (or expression) will each be denoted by $p(\cdot)$, as in (1.1) below, for example. Although an abuse of standard mathematical notation, this method is compact and similar to the standard practice of using $p(\cdot)$ for the probability of any discrete event, where the sample space is also suppressed in the notation. Depending on context, to avoid confusion, we may use the notation $\Pr(\cdot)$ for the probability of an event; for example, $\Pr(\theta > 2) = \int_{\theta>2} p(\theta)d\theta$. When using a standard distribution, we use a notation based on the name of the distribution; for example, if θ has a normal distribution with mean μ and variance σ^2, we write $\theta \sim \mathrm{N}(\mu, \sigma^2)$ or $p(\theta) = \mathrm{N}(\theta|\mu, \sigma^2)$ or, to be even more explicit, $p(\theta|\mu, \sigma^2) = \mathrm{N}(\theta|\mu, \sigma^2)$. Throughout, we use notation such as $\mathrm{N}(\mu, \sigma^2)$ for random variables and $\mathrm{N}(\theta|\mu, \sigma^2)$ for density functions. Notation and formulas for several standard distributions appear in Appendix A.

We also occasionally use the following expressions for all-positive random variables θ: the *coefficient of variation* is defined as $\mathrm{sd}(\theta)/\mathrm{E}(\theta)$, the *geometric mean* is $\exp(\mathrm{E}[\log(\theta)])$, and the *geometric standard deviation* is $\exp(\mathrm{sd}[\log(\theta)])$.

Bayes' rule

In order to make probability statements about θ given y, we must begin with a *model* providing a *joint probability distribution* for θ and y. The joint probability mass or density function can be written as a product of two densities that are often referred to as the *prior distribution* $p(\theta)$ and the *sampling distribution* (or *data distribution*) $p(y|\theta)$, respectively:

$$p(\theta, y) = p(\theta)p(y|\theta).$$

Simply conditioning on the known value of the data y, using the basic property of conditional probability known as Bayes' rule, yields the *posterior* density:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}, \tag{1.1}$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$, and the sum is over all possible values of θ (or $p(y) = \int p(\theta)p(y|\theta)d\theta$ in the case of continuous θ). An equivalent form of (1.1) omits the factor $p(y)$, which does not depend on θ and, with fixed y, can thus be considered a constant, yielding the *unnormalized posterior density*, which is the right side of (1.2):

$$p(\theta|y) \propto p(\theta)p(y|\theta). \tag{1.2}$$

The second term in this expression, $p(y|\theta)$, is taken here as a function of θ, not of y. These simple formulas encapsulate the technical core of Bayesian inference: the primary task of any specific application is to develop the model $p(\theta, y)$ and perform the computations to summarize $p(\theta|y)$ in appropriate ways.

Prediction

To make inferences about an unknown observable, often called predictive inferences, we follow a similar logic. Before the data y are considered, the distribution of the unknown but observable y is

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta. \tag{1.3}$$

This is often called the marginal distribution of y, but a more informative name is the *prior predictive distribution*: prior because it is not conditional on a previous observation of the process, and predictive because it is the distribution for a quantity that is observable.

After the data y have been observed, we can predict an unknown observable, $\tilde{y}$, from the same process. For example, $y = (y_1, \ldots, y_n)$ may be the vector of recorded weights of an object weighed n times on a scale, $\theta = (\mu, \sigma^2)$ may be the unknown true weight of the object and the measurement variance of the scale, and $\tilde{y}$ may be the yet to be recorded weight of the object in a planned new weighing. The distribution of $\tilde{y}$ is called the *posterior predictive distribution*, posterior because it is conditional on the observed y and predictive because it is a prediction for an observable $\tilde{y}$:

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta. \end{aligned} \tag{1.4}$$

The second and third lines display the posterior predictive distribution as an average of conditional predictions over the posterior distribution of θ. The last step follows from the assumed conditional independence of y and $\tilde{y}$ given θ.

Likelihood

Using Bayes' rule with a chosen probability model means that the data y affect the posterior inference (1.2) *only* through $p(y|\theta)$, which, when regarded as a function of θ, for fixed y, is called the *likelihood function*. In this way Bayesian inference obeys what is sometimes called the *likelihood principle*, which states that for a given sample of data, any two probability models $p(y|\theta)$ that have the same likelihood function yield the same inference for θ.

The likelihood principle is reasonable, but only within the framework of the model or family of models adopted for a particular analysis. In practice, one can rarely be confident that the chosen model is correct. We shall see in Chapter 6 that sampling distributions (imagining repeated realizations of our data) can play an important role in checking model assumptions. In fact, our view of an applied Bayesian statistician is one who is willing to apply Bayes' rule under a variety of possible models.

Likelihood and odds ratios

The ratio of the posterior density $p(\theta|y)$ evaluated at the points θ_1 and θ_2 under a given model is called the posterior *odds* for θ_1 compared to θ_2. The most familiar application of this concept is with discrete parameters, with θ_2 taken to be the complement of θ_1. Odds provide an alternative representation of probabilities and have the attractive property that Bayes' rule takes a particularly simple form when expressed in terms of them:

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)/p(y)}{p(\theta_2)p(y|\theta_2)/p(y)} = \frac{p(\theta_1)}{p(\theta_2)}\frac{p(y|\theta_1)}{p(y|\theta_2)}. \tag{1.5}$$

In words, the posterior odds are equal to the prior odds multiplied by the *likelihood ratio*, $p(y|\theta_1)/p(y|\theta_2)$.

1.4 Discrete examples: genetics and spell checking

We next demonstrate Bayes' theorem with two examples in which the immediate goal is inference about a particular discrete quantity rather than with the estimation of a parameter that describes an entire population. These discrete examples allow us to see the prior, likelihood, and posterior probabilities directly.

Inference about a genetic status

Human males have one X-chromosome and one Y-chromosome, whereas females have two X-chromosomes, each chromosome being inherited from one parent. Hemophilia is a disease that exhibits X-chromosome-linked recessive inheritance, meaning that a male who inherits the gene that causes the disease on the X-chromosome is affected, whereas a female carrying the gene on only one of her two X-chromosomes is not affected. The disease is generally fatal for women who inherit two such genes, and this is rare, since the frequency of occurrence of the gene is low in human populations.

Prior distribution. Consider a woman who has an affected brother, which implies that her mother must be a carrier of the hemophilia gene with one 'good' and one 'bad' hemophilia gene. We are also told that her father is not affected; thus the woman herself has a fifty-fifty chance of having the gene. The unknown quantity of interest, the state of the woman, has just two values: the woman is either a carrier of the gene ($\theta = 1$) or not ($\theta = 0$). Based on the information provided thus far, the prior distribution for the unknown θ can be expressed simply as $\Pr(\theta = 1) = \Pr(\theta = 0) = \frac{1}{2}$.

Data model and likelihood. The data used to update the prior information consist of the affection status of the woman's sons. Suppose she has two sons, neither of whom is affected. Let $y_i = 1$ or 0 denote an affected or unaffected son, respectively. The outcomes of the two sons are exchangeable and, conditional on the unknown θ, are independent; we assume the sons are not identical twins. The two items of independent data generate the following likelihood function:

$$\begin{aligned}\Pr(y_1=0, y_2=0 \mid \theta=1) &= (0.5)(0.5) = 0.25\\ \Pr(y_1=0, y_2=0 \mid \theta=0) &= (1)(1) = 1.\end{aligned}$$

These expressions follow from the fact that if the woman is a carrier, then each of her sons will have a 50% chance of inheriting the gene and so being affected, whereas if she is not a carrier then there is a probability close to 1 that a son of hers will be unaffected. (In fact, there is a nonzero probability of being affected even if the mother is not a carrier, but this risk—the mutation rate—is small and can be ignored for this example.)

Posterior distribution. Bayes' rule can now be used to combine the information in the data with the prior probability; in particular, interest is likely to focus on the posterior probability that the woman is a carrier. Using y to denote the joint data (y_1, y_2), this is simply

$$\begin{aligned}\Pr(\theta=1|y) &= \frac{p(y|\theta=1)\Pr(\theta=1)}{p(y|\theta=1)\Pr(\theta=1)+p(y|\theta=0)\Pr(\theta=0)}\\ &= \frac{(0.25)(0.5)}{(0.25)(0.5)+(1.0)(0.5)} = \frac{0.125}{0.625} = 0.20.\end{aligned}$$

Intuitively it is clear that if a woman has unaffected children, it is less probable that she is a carrier, and Bayes' rule provides a formal mechanism for determining the extent of the correction. The results can also be described in terms of prior and posterior odds. The prior odds of the woman being a carrier are $0.5/0.5 = 1$. The likelihood ratio based on the information about her two unaffected sons is $0.25/1 = 0.25$, so the posterior odds are $1 \cdot 0.25 = 0.25$. Converting back to a probability, we obtain $0.25/(1+0.25) = 0.2$, as before.

Adding more data. A key aspect of Bayesian analysis is the ease with which sequential analyses can be performed. For example, suppose that the woman has a third son, who is also unaffected. The entire calculation does not need to be redone; rather we use the previous posterior distribution as the new prior distribution, to obtain:

$$\Pr(\theta=1|y_1,y_2,y_3) = \frac{(0.5)(0.20)}{(0.5)(0.20)+(1)(0.8)} = 0.111.$$

Alternatively, if we suppose that the third son is affected, it is easy to check that the posterior probability of the woman being a carrier becomes 1 (again ignoring the possibility of a mutation).

Spelling correction

Classification of words is a problem of managing uncertainty. For example, suppose someone types 'radom.' How should that be read? It could be a misspelling or mistyping of 'random' or 'radon' or some other alternative, or it could be the intentional typing of 'radom' (as in its first use in this paragraph). What is the probability that 'radom' actually means random? If we label y as the data and θ as the word that the person was intending to type, then

$$\Pr(\theta \mid y=\text{`radom'}) \propto p(\theta)\Pr(y=\text{`radom'} \mid \theta). \tag{1.6}$$

This product is the unnormalized posterior density. In this case, if for simplicity we consider only three possibilities for the intended word, θ (random, radon, or radom), we can compute the posterior probability of interest by first computing the unnormalized density for all three values of theta and then normalizing:

$$p(\text{random}|\text{`radom'}) = \frac{p(\theta_1)p(\text{`radom'}|\theta_1)}{\sum_{j=1}^{3} p(\theta_j)p(\text{`radom'}|\theta_j)},$$

where θ_1=random, θ_2=radon, and θ_3=radom. The prior probabilities $p(\theta_j)$ can most simply come from frequencies of these words in some large database, ideally one that is adapted to the problem at hand (for example, a database of recent student emails if the word in question is appearing in such a document). The likelihoods $p(y|\theta_j)$ can come from some modeling of spelling and typing errors, perhaps fit using some study in which people were followed up after writing emails to identify any questionable words.

Prior distribution. Without any other context, it makes sense to assign the prior probabilities $p(\theta_j)$ based on the relative frequencies of these three words in some databases. Here are probabilities supplied by researchers at Google:

θ	$p(\theta)$
random	7.60×10^{-5}
radon	6.05×10^{-6}
radom	3.12×10^{-7}

Since we are considering only these possibilities, we could renormalize the three numbers to sum to 1 ($p(\text{random}) = \frac{760}{760+60.5+3.12}$, etc.) but there is no need, as the adjustment would merely be absorbed into the proportionality constant in (1.6).

Returning to the table above, we were surprised to see the probability of 'radom' in the corpus being as high as it was. We looked up the word in Wikipedia and found that it is a medium-sized city: home to 'the largest and best-attended air show in Poland ... also the popular unofficial name for a semiautomatic 9 mm Para pistol of Polish design ... ' For the documents that we encounter, the relative probability of 'radom' seems much too high. If the probabilities above do not seem appropriate for our application, this implies that we have prior information or beliefs that have not yet been included in the model. We shall return to this point after first working out the model's implications for this example.

Likelihood. Here are some conditional probabilities from Google's model of spelling and typing errors:

θ	$p(\text{'radom'}|\theta)$
random	0.00193
radon	0.000143
radom	0.975

We emphasize that this likelihood function is *not* a probability distribution. Rather, it is a set of conditional probabilities of a particular outcome ('radom') from three different probability distributions, corresponding to three different possibilities for the unknown parameter θ.

These particular values look reasonable enough—a 97% chance that this particular five-letter word will be typed correctly, a 0.2% chance of obtaining this character string by mistakenly dropping a letter from 'random,' and a much lower chance of obtaining it by mistyping the final letter of 'radon.' We have no strong intuition about these probabilities and will trust the Google engineers here.

Posterior distribution. We multiply the prior probability and the likelihood to get joint probabilities and then renormalize to get posterior probabilities:

θ	$p(\theta)p(\text{'radom'}|\theta)$	$p(\theta|\text{'radom'})$
random	1.47×10^{-7}	0.325
radon	8.65×10^{-10}	0.002
radom	3.04×10^{-7}	0.673

Thus, conditional on the model, the typed word 'radom' is about twice as likely to be correct as to be a typographical error for 'random,' and it is very unlikely to be a mistaken instance of 'radon.' A fuller analysis would include possibilities beyond these three words, but the basic idea is the same.

Decision making, model checking, and model improvement. We can envision two directions to go from here. The first approach is to accept the two-thirds probability that the word was typed correctly or even to simply declare 'radom' as correct on first pass. The second option would be to question this probability by saying, for example, that 'radom' looks like a typo and that the estimated probability of it being correct seems much too high.

When we dispute the claims of a posterior distribution, we are saying that the model does not fit the data or that we have additional prior information not included in the model so far. In this case, we are only examining one word so lack of fit is not the issue; thus a dispute over the posterior must correspond to a claim of additional information, either in the prior or the likelihood.

For this problem we have no particular grounds on which to criticize the likelihood. The prior probabilities, on the other hand, are highly context dependent. The word 'random' is of course highly frequent in our own writing on statistics, 'radon' occurs occasionally (see Section 9.4), while 'radom' was entirely new to us. Our surprise at the high probability of 'radom' represents additional knowledge relevant to our particular problem.

The model can be elaborated most immediately by including contextual information in the prior probabilities. For example, if the document under study is a statistics book, then it becomes more likely that the person intended to type 'random.' If we label x as the contextual information used by the model, the Bayesian calculation then becomes,

$$p(\theta|x, y) \propto p(\theta|x)p(y|\theta, x).$$

To first approximation, we can simplify that last term to $p(y|\theta)$, so that the probability of any particular error (that is, the probability of typing a particular string y given the intended word θ) does not depend on context. This is not a perfect assumption but could reduce the burden of modeling and computation.

The practical challenges in Bayesian inference involve setting up models to estimate all these probabilities from data. At that point, as shown above, Bayes' rule can be easily applied to determine the implications of the model for the problem at hand.

1.5 Probability as a measure of uncertainty

We have already used concepts such as probability density, and indeed we assume that the reader has a fair degree of familiarity with basic probability theory (although in Section 1.8 we provide a brief technical review of some probability calculations that often arise in Bayesian analysis). But since the uses of probability within a Bayesian framework are much broader than within non-Bayesian statistics, it is important to consider at least briefly the foundations of the concept of probability before considering more detailed statistical examples. We take for granted a common understanding on the part of the reader of the mathematical definition of probability: that probabilities are numerical quantities, defined on a set of 'outcomes,' that are nonnegative, additive over mutually exclusive outcomes, and sum to 1 over all possible mutually exclusive outcomes.

In Bayesian statistics, probability is used as the fundamental measure or yardstick of uncertainty. Within this paradigm, it is equally legitimate to discuss the probability of 'rain tomorrow' or of a Brazilian victory in the soccer World Cup as it is to discuss the probability that a coin toss will land heads. Hence, it becomes as natural to consider the probability that an unknown estimand lies in a particular range of values as it is to consider the probability that the mean of a random sample of 10 items from a known fixed population of size 100 will lie in a certain range. The first of these two probabilities is of more interest after data have been acquired whereas the second is more relevant beforehand. Bayesian methods enable statements to be made about the partial knowledge available (based on data) concerning some situation or 'state of nature' (unobservable or as yet unobserved) in a systematic way, using probability as the yardstick. The guiding principle is that the state of knowledge about anything unknown is described by a probability distribution.

What is meant by a numerical measure of uncertainty? For example, the probability of 'heads' in a coin toss is widely agreed to be $\frac{1}{2}$. Why is this so? Two justifications seem to be commonly given:

1. Symmetry or exchangeability argument:

$$\text{probability} = \frac{\text{number of favorable cases}}{\text{number of possibilities}},$$

   assuming equally likely possibilities. For a coin toss this is really a physical argument, based on assumptions about the forces at work in determining the manner in which the coin will fall, as well as the initial physical conditions of the toss.

2. Frequency argument: probability = relative frequency obtained in a long sequence of tosses, assumed to be performed in an identical manner, physically independently of each other.

Both the above arguments are in a sense subjective, in that they require judgments about the nature of the coin and the tossing procedure, and both involve semantic arguments about the meaning of equally likely events, identical measurements, and independence. The frequency argument may be perceived to have certain special difficulties, in that it involves the hypothetical notion of a long sequence of identical tosses. If taken strictly, this point of view does not allow a statement of probability for a single coin toss that does not happen to be embedded, at least conceptually, in a long sequence of identical events.

The following examples illustrate how probability judgments can be increasingly subjective. First, consider the following modified coin experiment. Suppose that a particular coin is stated to be either double-headed *or* double-tailed, with no further information provided. Can one still talk of the probability of heads? It seems clear that in common parlance one certainly can. It is less clear, perhaps, how to assess this new probability, but many would agree on the same value of $\frac{1}{2}$, perhaps based on the exchangeability of the labels 'heads' and 'tails.'

Now consider some further examples. Suppose Colombia plays Brazil in soccer tomorrow: what is the probability of Colombia winning? What is the probability of rain tomorrow? What is the probability that Colombia wins, if it rains tomorrow? What is the probability that a specified rocket launch will fail? Although each of these questions seems reasonable in a common-sense way, it is difficult to contemplate strong frequency interpretations for the probabilities being referenced. Frequency interpretations can usually be *constructed*, however, and this is an extremely useful tool in statistics. For example, one can consider the future rocket launch as a sample from the population of potential launches of the same type, and look at the frequency of past launches that have failed (see the bibliographic note at the end of this chapter for more details on this example). Doing this sort of thing scientifically means creating a probability model (or, at least, a 'reference set' of comparable events), and this brings us back to a situation analogous to the simple coin toss, where we must consider the outcomes in question as exchangeable and thus equally likely.

Why is probability a reasonable way of quantifying uncertainty? The following reasons are often advanced.

1. By analogy: physical randomness induces uncertainty, so it seems reasonable to describe uncertainty in the language of random events. Common speech uses many terms such as 'probably' and 'unlikely,' and it appears consistent with such usage to extend a more formal probability calculus to problems of scientific inference.

2. Axiomatic or normative approach: related to decision theory, this approach places all statistical inference in the context of decision-making with gains and losses. Then reasonable

axioms (ordering, transitivity, and so on) imply that uncertainty *must* be represented in terms of probability. We view this normative rationale as suggestive but not compelling.

3. Coherence of bets. *Define* the probability p attached (by you) to an event E as the fraction ($p \in [0, 1]$) at which you would exchange (that is, bet) \$$p$ for a return of \$1 if E occurs. That is, if E occurs, you gain \$$(1 - p)$; if the complement of E occurs, you lose \$$p$. For example:
   - Coin toss: thinking of the coin toss as a fair bet suggests even odds corresponding to $p = \frac{1}{2}$.
   - Odds for a game: if you are willing to bet on team A to win a game at 10 to 1 odds against team B (that is, you bet 1 to win 10), your 'probability' for team A winning is at least $\frac{1}{11}$.

   The principle of coherence states that your assignment of probabilities to all possible events should be such that it is not possible to make a definite gain by betting with you. It can be proved that probabilities constructed under this principle must satisfy the basic axioms of probability theory.

   The betting rationale has some fundamental difficulties:
   - Exact odds are required, on which you would be willing to bet in either direction, for all events. How can you assign exact odds if you are not sure?
   - If a person is willing to bet with you, and has information you do not, it might not be wise for you to take the bet. In practice, probability is an incomplete (necessary but not sufficient) guide to betting.

All of these considerations suggest that probabilities may be a reasonable approach to summarizing uncertainty in applied statistics, but the ultimate proof is in the success of the applications. The remaining chapters of this book demonstrate that probability provides a rich and flexible framework for handling uncertainty in statistical applications.

Subjectivity and objectivity

All statistical methods that use probability are subjective in the sense of relying on mathematical idealizations of the world. Bayesian methods are sometimes said to be especially subjective because of their reliance on a prior distribution, but in most problems, scientific judgment is necessary to specify both the 'likelihood' and the 'prior' parts of the model. For example, linear regression models are generally at least as suspect as any prior distribution that might be assumed about the regression parameters. A general principle is at work here: whenever there is replication, in the sense of many exchangeable units observed, there is scope for estimating features of a probability distribution from data and thus making the analysis more 'objective.' If an experiment as a whole is replicated several times, then the parameters of the prior distribution can themselves be estimated from data, as discussed in Chapter 5. In any case, however, certain elements requiring scientific judgment will remain, notably the choice of data included in the analysis, the parametric forms assumed for the distributions, and the ways in which the model is checked.

1.6 Example: probabilities from football point spreads

As an example of how probabilities might be assigned using empirical data and plausible substantive assumptions, we consider methods of estimating the probabilities of certain outcomes in professional (American) football games. This is an example only of probability assignment, not of Bayesian inference. A number of approaches to assigning probabilities for football game outcomes are illustrated: making subjective assessments, using empirical probabilities based on observed data, and constructing a parametric probability model.

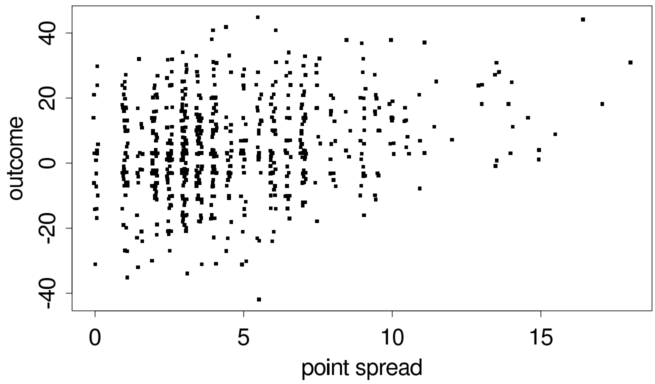


Figure 1.1 *Scatterplot of actual outcome vs. point spread for each of 672 professional football games. The x and y coordinates are jittered by adding uniform random numbers to each point's coordinates (between -0.1 and 0.1 for the x coordinate; between -0.2 and 0.2 for the y coordinate) in order to display multiple values but preserve the discrete-valued nature of each.*

Football point spreads and game outcomes

Football experts provide a *point spread* for every football game as a measure of the difference in ability between the two teams. For example, team A might be a 3.5-point favorite to defeat team B. The implication of this point spread is that the proposition that team A, the favorite, defeats team B, the underdog, by 4 or more points is considered a fair bet; in other words, the probability that A wins by more than 3.5 points is $\frac{1}{2}$. If the point spread is an integer, then the implication is that team A is as likely to win by more points than the point spread as it is to win by fewer points than the point spread (or to lose); there is positive probability that A will win by exactly the point spread, in which case neither side is paid off. The assignment of point spreads is itself an interesting exercise in probabilistic reasoning; one interpretation is that the point spread is the median of the distribution of the gambling population's beliefs about the possible outcomes of the game. For the rest of this example, we treat point spreads as given and do not worry about how they were derived.

The point spread and actual game outcome for 672 professional football games played during the 1981, 1983, and 1984 seasons are graphed in Figure 1.1. (Much of the 1982 season was canceled due to a labor dispute.) Each point in the scatterplot displays the point spread, x, and the actual outcome (favorite's score minus underdog's score), y. (In games with a point spread of zero, the labels 'favorite' and 'underdog' were assigned at random.) A small random jitter is added to the x and y coordinate of each point on the graph so that multiple points do not fall exactly on top of each other.

Assigning probabilities based on observed frequencies

It is of interest to assign probabilities to particular events: Pr(favorite wins), Pr(favorite wins | point spread is 3.5 points), Pr(favorite wins by more than the point spread), Pr(favorite wins by more than the point spread | point spread is 3.5 points), and so forth. We might report a subjective probability based on informal experience gathered by reading the newspaper and watching football games. The probability that the favored team wins a game should certainly be greater than 0.5, perhaps between 0.6 and 0.75? More complex events require more intuition or knowledge on our part. A more systematic approach is to assign probabilities based on the data in Figure 1.1. Counting a tied game as one-half win and

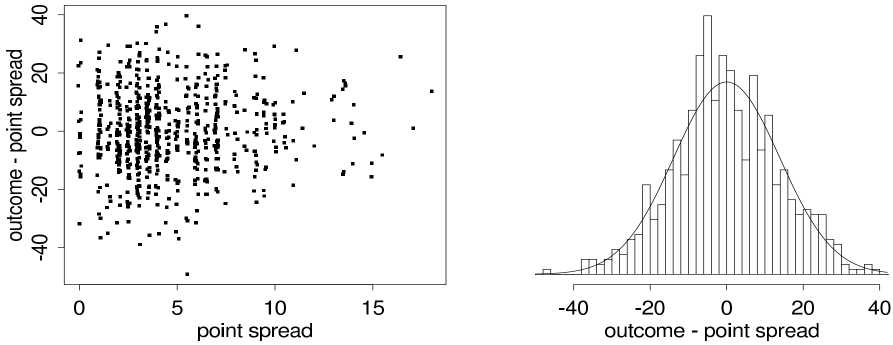


Figure 1.2 *(a) Scatterplot of (actual outcome − point spread) vs. point spread for each of 672 professional football games (with uniform random jitter added to x and y coordinates). (b) Histogram of the differences between the game outcome and the point spread, with the* $N(0, 14^2)$ *density superimposed.*

one-half loss, and ignoring games for which the point spread is zero (and thus there is no favorite), we obtain empirical estimates such as:

- Pr(favorite wins) $= \frac{410.5}{655} = 0.63$
- Pr(favorite wins $| \, x = 3.5) = \frac{36}{59} = 0.61$
- Pr(favorite wins by more than the point spread) $= \frac{308}{655} = 0.47$
- Pr(favorite wins by more than the point spread $| \, x = 3.5) = \frac{32}{59} = 0.54$.

These empirical probability assignments all seem sensible in that they match the intuition of knowledgeable football fans. However, such probability assignments are problematic for events with few directly relevant data points. For example, 8.5-point favorites won five out of five times during this three-year period, whereas 9-point favorites won thirteen out of twenty times. However, we realistically expect the probability of winning to be greater for a 9-point favorite than for an 8.5-point favorite. The small sample size with point spread 8.5 leads to imprecise probability assignments. We consider an alternative method using a parametric model.

A parametric model for the difference between outcome and point spread

Figure 1.2a displays the differences $y-x$ between the observed game outcome and the point spread, plotted versus the point spread, for the games in the football dataset. (Once again, random jitter was added to both coordinates.) This plot suggests that it may be roughly reasonable to model the distribution of $y-x$ as independent of x. (See Exercise 6.10.) Figure 1.2b is a histogram of the differences $y-x$ for all the football games, with a fitted normal density superimposed. This plot suggests that it may be reasonable to approximate the marginal distribution of the random variable $d = y - x$ by a normal distribution. The sample mean of the 672 values of d is 0.07, and the sample standard deviation is 13.86, suggesting that the results of football games are approximately normal with mean equal to the point spread and standard deviation nearly 14 points (two converted touchdowns). For the remainder of the discussion we take the distribution of d to be independent of x and normal with mean zero and standard deviation 14 for each x; that is,

$$d|x \sim \mathrm{N}(0, 14^2),$$

as displayed in Figure 1.2b. The assigned probability model is not perfect: it does not fit the data exactly, and, as is often the case with real data, neither football scores nor point spreads are continuous-valued quantities.

Assigning probabilities using the parametric model

Nevertheless, the model provides a convenient approximation that can be used to assign probabilities to events. If d has a normal distribution with mean zero and is independent of the point spread, then the probability that the favorite wins by more than the point spread is $\frac{1}{2}$, conditional on any value of the point spread, and therefore unconditionally as well. Denoting probabilities obtained by the normal model as $\Pr_{\rm norm}$, the probability that an x-point favorite wins the game can be computed, assuming the normal model, as follows:

$$\Pr_{\rm norm}(y>0\,|\,x) = \Pr_{\rm norm}(d>-x\,|\,x) = 1-\Phi\left(-\frac{x}{14}\right),$$

where Φ is the standard normal cumulative distribution function. For example,

- $\Pr_{\rm norm}(\text{favorite wins}\,|\,x=3.5) = 0.60$
- $\Pr_{\rm norm}(\text{favorite wins}\,|\,x=8.5) = 0.73$
- $\Pr_{\rm norm}(\text{favorite wins}\,|\,x=9.0) = 0.74$.

The probability for a 3.5-point favorite agrees with the empirical value given earlier, whereas the probabilities for 8.5- and 9-point favorites make more intuitive sense than the empirical values based on small samples.

1.7 Example: calibration for record linkage

We emphasize the essentially empirical (not 'subjective' or 'personal') nature of probabilities with another example in which they are estimated from data.

Record linkage refers to the use of an algorithmic technique to identify records from different databases that correspond to the same individual. Record-linkage techniques are used in a variety of settings. The work described here was formulated and first applied in the context of record linkage between the U.S. Census and a large-scale post-enumeration survey, which is the first step of an extensive matching operation conducted to evaluate census coverage for subgroups of the population. The goal of this first step is to declare as many records as possible 'matched' by computer without an excessive rate of error, thereby avoiding the cost of the resulting manual processing for all records not declared 'matched.'

Existing methods for assigning scores to potential matches

Much attention has been paid in the record-linkage literature to the problem of assigning 'weights' to individual fields of information in a multivariate record and obtaining a composite 'score,' which we call y, that summarizes the closeness of agreement between two records. Here, we assume that this step is complete in the sense that these rules have been chosen. The next step is the assignment of candidate matched pairs, where each pair of records consists of the best potential match for each other from the respective databases. The specified weighting rules then order the candidate matched pairs. In the motivating problem at the Census Bureau, a binary choice is made between the alternatives 'declare matched' vs. 'send to followup,' where a cutoff score is needed above which records are declared matched. The false-match rate is then defined as the number of falsely matched pairs divided by the number of declared matched pairs.

Particularly relevant for any such decision problem is an accurate method for assessing the probability that a candidate matched pair is a correct match as a function of its score. Simple methods exist for converting the scores into probabilities, but these lead to extremely inaccurate, typically grossly optimistic, estimates of false-match rates. For example, a manual check of a set of records with nominal false-match probabilities ranging from 10^{-3} to 10^{-7} (that is, pairs deemed almost certain to be matches) found actual false-match rates

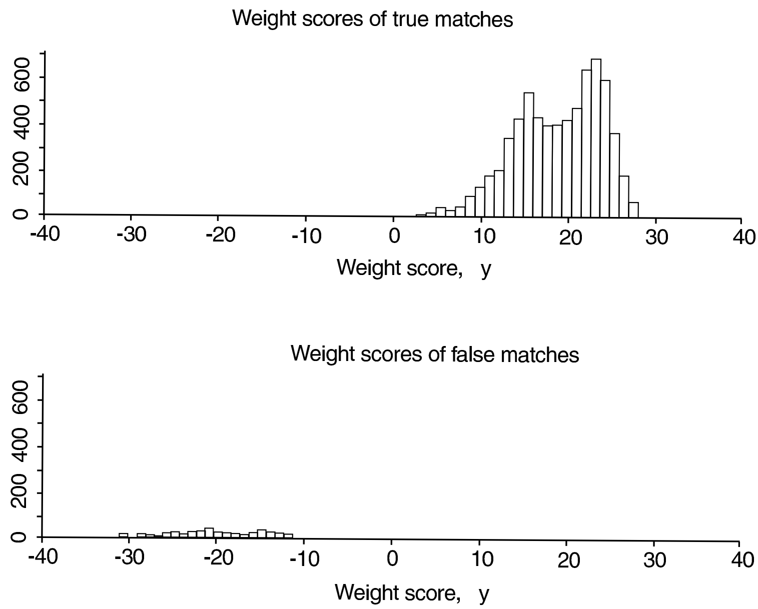


Figure 1.3 *Histograms of weight scores y for true and false matches in a sample of records from the 1988 test Census. Most of the matches in the sample are true (because a pre-screening process has already picked these as the best potential match for each case), and the two distributions are mostly, but not completely, separated.*

closer to the 1% range. Records with nominal false-match probabilities of 1% had an actual false-match rate of 5%.

We would like to use Bayesian methods to recalibrate these to obtain objective probabilities of matching for a given decision rule—in the same way that in the football example, we used past data to estimate the probabilities of different game outcomes conditional on the point spread. Our approach is to work with the scores y and empirically estimate the probability of a match as a function of y.

Estimating match probabilities empirically

We obtain accurate match probabilities using mixture modeling, a topic we discuss in detail in Chapter 22. The distribution of previously obtained scores for the candidate matches is considered a 'mixture' of a distribution of scores for true matches and a distribution for non-matches. The parameters of the mixture model are estimated from the data. The estimated parameters allow us to calculate an estimate of the probability of a false match (a pair declared matched that is not a true match) for any given decision threshold on the scores. In the procedure that was actually used, some elements of the mixture model (for example, the optimal transformation required to allow a mixture of normal distributions to apply) were fit using 'training' data with known match status (separate from the data to which we apply our calibration procedure), but we do not describe those details here. Instead we focus on how the method would be used with a set of data with unknown match status.

Support for this approach is provided in Figure 1.3, which displays the distribution of scores for the matches and non-matches in a particular dataset obtained from 2300 records from a 'test Census' survey conducted in a single local area two years before the 1990 Census. The two distributions, $p(y|\text{match})$ and $p(y|\text{non-match})$, are mostly distinct—meaning that

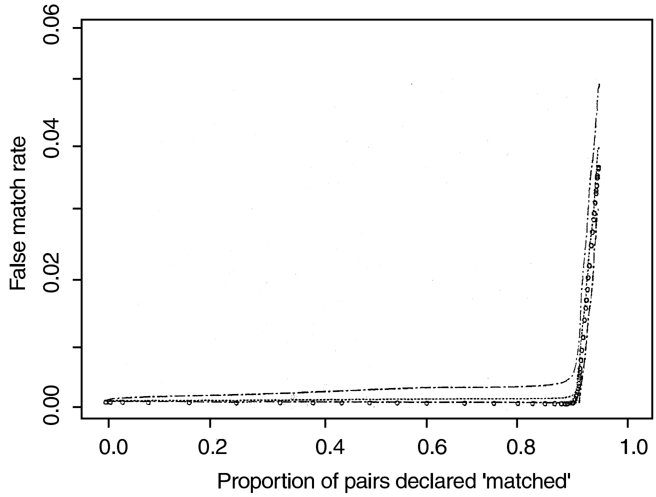


Figure 1.4 *Lines show expected false-match rate (and 95% bounds) as a function of the proportion of cases declared matches, based on the mixture model for record linkage. Dots show the actual false-match rate for the data.*

in most cases it is possible to identify a candidate as a match or not given the score alone—but with some overlap.

In our application dataset, we do not know the match status. Thus we are faced with a single combined histogram from which we estimate the two component distributions and the proportion of the population of scores that belong to each component. Under the mixture model, the distribution of scores can be written as,

$$p(y) = \Pr(\text{match})\, p(y|\text{match}) + \Pr(\text{non-match})\, p(y|\text{non-match}). \tag{1.7}$$

The mixture probability (Pr(match)) and the parameters of the distributions of matches ($p(y|\text{match})$) and non-matches ($p(y|\text{non-match})$) are estimated using the mixture model approach (as described in Chapter 22) applied to the combined histogram from the data with unknown match status.

To use the method to make record-linkage decisions, we construct a curve giving the false-match rate as a function of the decision threshold, the score above which pairs will be 'declared' a match. For a given decision threshold, the probability distributions in (1.7) can be used to estimate the probability of a false match, a score y above the threshold originating from the distribution $p(y|\text{non-match})$. The lower the threshold, the more pairs we will declare as matches. As we declare more matches, the proportion of errors increases. The approach described here should provide an objective error estimate for each threshold. (See the validation in the next paragraph.) Then a decision maker can determine the threshold that provides an acceptable balance between the goals of declaring more matches automatically (thus reducing the clerical labor) and making fewer mistakes.

External validation of the probabilities using test data

The approach described above was externally validated using data for which the match status is known. The method was applied to data from three different locations of the 1988 test Census, and so three tests of the methods were possible. We provide detailed results for one; results for the other two were similar. The mixture model was fitted to the scores of all the candidate pairs at a test site. Then the estimated model was used to create the lines in Figure 1.4, which show the expected false-match rate (and uncertainty bounds) in

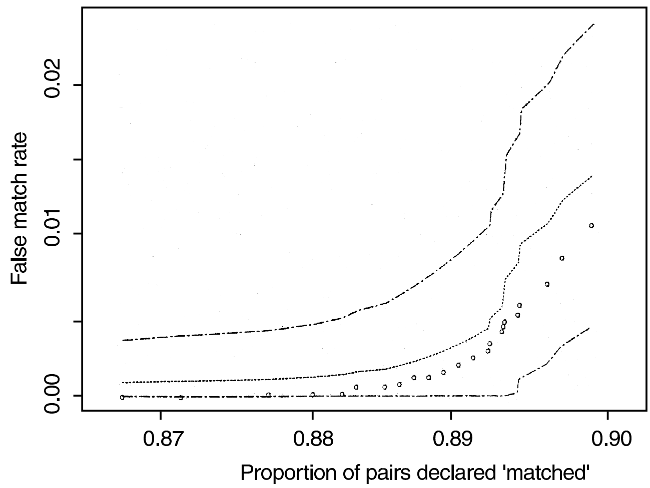


Figure 1.5 *Expansion of Figure 1.4 in the region where the estimated and actual match rates change rapidly. In this case, it would seem a good idea to match about 88% of the cases and send the rest to followup.*

terms of the proportion of cases declared matched, as the threshold varies from high (thus allowing no matches) to low (thus declaring almost all the candidate pairs to be matches). The false-match proportion is an increasing function of the number of declared matches, which makes sense: as we move rightward on the graph, we are declaring weaker and weaker cases to be matches.

The lines on Figure 1.4 display the expected proportion of false matches and 95% posterior bounds for the false-match rate as estimated from the model. (These bounds give the estimated range within which there is 95% posterior probability that the false-match rate lies. The concept of posterior intervals is discussed in more detail in the next chapter.) The dots in the graph display the actual false-match proportions, which track well with the model. In particular, the model would suggest a recommendation of declaring something less than 90% of cases as matched and giving up on the other 10% or so, so as to avoid most of the false matches, and the dots show a similar pattern.

It is clearly possible to match large proportions of the files with little or no error. Also, the quality of candidate matches becomes dramatically worse at some point where the false-match rate accelerates. Figure 1.5 takes a magnifying glass to the previous display to highlight the behavior of the calibration procedure in the region of interest where the false-match rate accelerates. The predicted false-match rate curves bend upward, close to the points where the observed false-match rate curves rise steeply, which is a particularly encouraging feature of the calibration method. The calibration procedure performs well from the standpoint of providing predicted probabilities that are close to the true probabilities and interval estimates that are informative and include the true values. By comparison, the original estimates of match probabilities, constructed by multiplying weights without empirical calibration, were highly inaccurate.

1.8 Some useful results from probability theory

We assume the reader is familiar with elementary manipulations involving probabilities and probability distributions. In particular, basic probability background that must be well understood for key parts of the book includes the manipulation of joint densities, the definition of simple moments, the transformation of variables, and methods of simulation. In this section we briefly review these assumed prerequisites and clarify some further notational conventions used in the remainder of the book. Appendix A provides information on some commonly used probability distributions.

As introduced in Section 1.3, we generally represent joint distributions by their joint probability mass or density function, with dummy arguments reflecting the name given to each variable being considered. Thus for two quantities u and v, we write the joint density as $p(u, v)$; if specific values need to be referenced, this notation will be further abused as with, for example, $p(u, v=1)$.

In Bayesian calculations relating to a joint density $p(u, v)$, we will often refer to a *conditional* distribution or density function such as $p(u|v)$ and a *marginal* density such as $p(u) = \int p(u, v)dv$. In this notation, either or both u and v can be vectors. Typically it will be clear from the context that the range of integration in the latter expression refers to the entire range of the variable being integrated out. It is also often useful to *factor* a joint density as a product of marginal and conditional densities; for example, $p(u, v, w) = p(u|v, w)p(v|w)p(w)$.

Some authors use different notations for distributions on parameters and observables—for example, $\pi(\theta), f(y|\theta)$—but this obscures the fact that all probability distributions have the same *logical* status in Bayesian inference. We must always be careful, though, to indicate appropriate conditioning; for example, $p(y|\theta)$ is different from $p(y)$. In the interests of conciseness, however, our notation hides the conditioning on hypotheses that hold throughout—no probability judgments can be made in a vacuum—and to be more explicit one might use a notation such as the following:

$$p(\theta, y|H) = p(\theta|H)p(y|\theta, H),$$

where H refers to the set of hypotheses or assumptions used to define the model. Also, we sometimes suppress explicit conditioning on known explanatory variables, x.

We use the standard notations, $\mathrm{E}(\cdot)$ and $\mathrm{var}(\cdot)$, for mean and variance, respectively:

$$\mathrm{E}(u) = \int up(u)du, \quad \mathrm{var}(u) = \int (u - \mathrm{E}(u))^2 p(u)du.$$

For a vector parameter u, the expression for the mean is the same, and the covariance matrix is defined as

$$\mathrm{var}(u) = \int (u - \mathrm{E}(u))(u - \mathrm{E}(u))^T p(u)du,$$

where u is considered a column vector. (We use the terms 'variance matrix' and 'covariance matrix' interchangeably.) This notation is slightly imprecise, because $\mathrm{E}(u)$ and $\mathrm{var}(u)$ are really functions of the distribution function, $p(u)$, not of the variable u. In an expression involving an expectation, any variable that does not appear explicitly as a conditioning variable is assumed to be integrated out in the expectation; for example, $\mathrm{E}(u|v)$ refers to the conditional expectation of u with v held fixed—that is, the conditional expectation as a function of v—whereas $\mathrm{E}(u)$ is the expectation of u, averaging over v (as well as u).

Modeling using conditional probability

Useful probability models often express the distribution of observables conditionally or hierarchically rather than through more complicated unconditional distributions. For example, suppose y is the height of a university student selected at random. The marginal distribution $p(y)$ is (essentially) a mixture of two approximately normal distributions centered around 160 and 175 centimeters. A more useful description of the distribution of y would be based on the joint distribution of height and sex: $p(\text{male}) \approx p(\text{female}) \approx \frac{1}{2}$, along with the conditional specifications that $p(y|\text{female})$ and $p(y|\text{male})$ are each approximately normal with means 160 and 175 cm, respectively. If the conditional variances are not too large, the marginal distribution of y is bimodal. In general, we prefer to model complexity with a hierarchical structure using additional variables rather than with complicated marginal distributions, even when the additional variables are unobserved or even unobservable; this theme underlies mixture models, as discussed in Chapter 22. We repeatedly return to the theme of conditional modeling throughout the book.

Means and variances of conditional distributions

It is often useful to express the mean and variance of a random variable u in terms of the conditional mean and variance given some related quantity v. The mean of u can be obtained by averaging the conditional mean over the marginal distribution of v,

$$\mathrm{E}(u) = \mathrm{E}(\mathrm{E}(u|v)), \tag{1.8}$$

where the inner expectation averages over u, conditional on v, and the outer expectation averages over v. Identity (1.8) is easy to derive by writing the expectation in terms of the joint distribution of u and v and then factoring the joint distribution:

$$\mathrm{E}(u) = \int\int up(u,v)dudv = \int\int u\,p(u|v)du\,p(v)dv = \int \mathrm{E}(u|v)p(v)dv.$$

The corresponding result for the variance includes two terms, the mean of the conditional variance and the variance of the conditional mean:

$$\mathrm{var}(u) = \mathrm{E}(\mathrm{var}(u|v)) + \mathrm{var}(\mathrm{E}(u|v)). \tag{1.9}$$

This result can be derived by expanding the terms on the right side of (1.9):

$$\begin{aligned}
\mathrm{E}\left(\mathrm{var}(u|v)\right) + \mathrm{var}\left(\mathrm{E}(u|v)\right) &= \mathrm{E}\left(\mathrm{E}(u^2|v) - (\mathrm{E}(u|v))^2\right) + \mathrm{E}\left((\mathrm{E}(u|v))^2\right) - \left(\mathrm{E}\left(\mathrm{E}(u|v)\right)\right)^2 \\
&= \mathrm{E}(u^2) - \mathrm{E}\left((\mathrm{E}(u|v))^2\right) + \mathrm{E}\left((\mathrm{E}(u|v))^2\right) - (\mathrm{E}(u))^2 \\
&= \mathrm{E}(u^2) - (\mathrm{E}(u))^2 \\
&= \mathrm{var}(u).
\end{aligned}$$

Identities (1.8) and (1.9) also hold if u is a vector, in which case $\mathrm{E}(u)$ is a vector and $\mathrm{var}(u)$ a matrix.

Transformation of variables

It is common to transform a probability distribution from one parameterization to another. We review the basic result here for a probability density on a transformed space. For clarity, we use subscripts here instead of our usual generic notation, $p(\cdot)$. Suppose $p_u(u)$ is the density of the vector u, and we transform to $v = f(u)$, where v has the same number of components as u.

If p_u is a discrete distribution, and f is a one-to-one function, then the density of v is given by

$$p_v(v) = p_u(f^{-1}(v)).$$

If f is a many-to-one function, then a sum of terms appears on the right side of this expression for $p_v(v)$, with one term corresponding to each of the branches of the inverse function.

If p_u is a continuous distribution, and $v = f(u)$ is a one-to-one transformation, then the joint density of the transformed vector is

$$p_v(v) = |J|\,p_u(f^{-1}(v))$$

where $|J|$ is the absolute value of the determinant of the Jacobian of the transformation $u = f^{-1}(v)$ as a function of v; the Jacobian J is the square matrix of partial derivatives (with dimension given by the number of components of u), with the (i, j)th entry equal to $\partial u_i/\partial v_j$. Once again, if f is many-to-one, then $p_v(v)$ is a sum or integral of terms.

In one dimension, we commonly use the logarithm to transform the parameter space from $(0, \infty)$ to $(-\infty, \infty)$. When working with parameters defined on the open unit interval, $(0, 1)$, we often use the logistic transformation:

$$\text{logit}(u) = \log\left(\frac{u}{1-u}\right), \tag{1.10}$$

whose inverse transformation is

$$\text{logit}^{-1}(v) = \frac{e^v}{1+e^v}.$$

Another common choice is the probit transformation, $\Phi^{-1}(u)$, where Φ is the standard normal cumulative distribution function, to transform from $(0, 1)$ to $(-\infty, \infty)$.

1.9 Computation and software

At the time of writing, the authors rely primarily on the software package R for graphs and basic simulations, fitting of classical simple models (including regression, generalized linear models, and nonparametric methods such as locally weighted regression), optimization, and some simple programming. We use the Bayesian inference package Stan (see Appendix C) for fitting most models, but for teaching purposes in this book we describe how to perform most of the computations from first principles. Even when using Stan, we typically work within R to plot and transform the data before model fitting, and to display inferences and model checks afterwards.

Specific computational tasks that arise in Bayesian data analysis include:

- Vector and matrix manipulations (see Table 1.1)
- Computing probability density functions (see Appendix A)
- Drawing simulations from probability distributions (see Appendix A for standard distributions and Exercise 1.9 for an example of a simple stochastic process)
- Structured programming (including looping and customized functions)
- Calculating the linear regression estimate and variance matrix (see Chapter 14)
- Graphics, including scatterplots with overlain lines and multiple graphs per page (see Chapter 6 for examples).

Our general approach to computation is to fit many models, gradually increasing the complexity. We do *not* recommend the strategy of writing a model and then letting the computer run overnight to estimate it perfectly. Rather, we prefer to fit each model relatively quickly, using inferences from the previously fitted simpler models as starting values, and displaying inferences and comparing to data before continuing.

We discuss computation in detail in Part III of this book after first introducing the fundamental concepts of Bayesian modeling, inference, and model checking. Appendix C illustrates how to perform computations in R and Stan in several different ways for a single example.

Summarizing inferences by simulation

Simulation forms a central part of much applied Bayesian analysis, because of the relative ease with which samples can often be generated from a probability distribution, even when the density function cannot be explicitly integrated. In performing simulations, it is helpful to consider the duality between a probability density function and a histogram of a set of random draws from the distribution: given a large enough sample, the histogram can provide practically complete information about the density, and in particular, various sample moments, percentiles, and other summary statistics provide estimates of any aspect of the distribution, to a level of precision that can be estimated. For example, to estimate the 95th percentile of the distribution of θ, draw a random sample of size S from $p(\theta)$ and use the $0.95S$th order statistic. For most purposes, $S = 1000$ is adequate for estimating the 95th percentile in this way.

Another advantage of simulation is that extremely large or small simulated values often flag a problem with model specification or parameterization (for example, see Figure 4.2) that might not be noticed if estimates and probability statements were obtained in analytic form.

Generating values from a probability distribution is often straightforward with modern computing techniques based on (pseudo)random number sequences. A well-designed pseudorandom number generator yields a deterministic sequence that appears to have the same properties as a sequence of independent random draws from the uniform distribution on $[0, 1]$. Appendix A describes methods for drawing random samples from some commonly used distributions.

Sampling using the inverse cumulative distribution function

As an introduction to the ideas of simulation, we describe a method for sampling from discrete and continuous distributions using the inverse cumulative distribution function. The *cumulative distribution function*, or *cdf*, F, of a one-dimensional distribution, $p(v)$, is defined by

$$\begin{aligned} F(v_*) &= \Pr(v \leq v_*) \\ &= \begin{cases} \sum_{v \leq v_*} p(v) & \text{if } p \text{ is discrete} \\ \int_{-\infty}^{v_*} p(v)dv & \text{if } p \text{ is continuous.} \end{cases} \end{aligned}$$

The inverse cdf can be used to obtain random samples from the distribution p, as follows. First draw a random value, U, from the uniform distribution on $[0, 1]$, using a table of random numbers or, more likely, a random number function on the computer. Now let $v = F^{-1}(U)$. The function F is not necessarily one-to-one—certainly not if the distribution is discrete—but $F^{-1}(U)$ is unique with probability 1. The value v will be a random draw from p, and is easy to compute as long as $F^{-1}(U)$ is simple. For a discrete distribution, F^{-1} can simply be tabulated.

For a continuous example, suppose v has an exponential distribution with parameter λ (see Appendix A); then its cdf is $F(v) = 1 - e^{-\lambda v}$, and the value of v for which $U = F(v)$ is $v = -\frac{\log(1-U)}{\lambda}$. Then, recognizing that $1 - U$ also has the uniform distribution on $[0, 1]$, we see we can obtain random draws from the exponential distribution as $-\frac{\log U}{\lambda}$. We discuss other methods of simulation in Part III of the book and Appendix A.

Simulation of posterior and posterior predictive quantities

In practice, we are most often interested in simulating draws from the posterior distribution of the model parameters θ, and perhaps from the posterior predictive distribution of unknown observables $\tilde{y}$. Results from a set of S simulation draws can be stored in the computer in an array, as illustrated in Table 1.1. We use the notation $s = 1, \ldots, S$ to index simulation draws; $(\theta^s, \tilde{y}^s)$ is the corresponding joint draw of parameters and predicted quantities from their joint posterior distribution.

Simulation draw	Parameters			Predictive quantities		
	θ_1	$\dots$	θ_k	$\tilde{y}_1$	$\dots$	$\tilde{y}_n$
1	θ_1^1	$\dots$	θ_k^1	$\tilde{y}_1^1$	$\dots$	$\tilde{y}_n^1$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
S	θ_1^S	$\dots$	θ_k^S	$\tilde{y}_1^S$	$\dots$	$\tilde{y}_n^S$

Table 1.1 *Structure of posterior and posterior predictive simulations. The superscripts are indexes, not powers.*

From these simulated values, we can estimate the posterior distribution of any quantity of interest, such as θ_1/θ_3, by just computing a new column in Table 1.1 using the existing S draws of $(\theta, \tilde{y})$. We can estimate the posterior probability of any event, such as $\Pr(\tilde{y}_1 + \tilde{y}_2 > e^{\theta_1})$, by the proportion of the S simulations for which it is true. We are often interested in posterior intervals; for example, the central 95% posterior interval $[a, b]$ for the parameter θ_j, for which $\Pr(\theta_j < a) = 0.025$ and $\Pr(\theta_j > b) = 0.025$. These values can be directly estimated by the appropriate simulated values of θ_j, for example, the 25th and 976th order statistics if $S = 1000$. We commonly summarize inferences by 50% and 95% intervals.

We return to the accuracy of simulation inferences in Section 10.5 after we have gained some experience using simulations of posterior distributions in some simple examples.

1.10 Bayesian inference in applied statistics

A pragmatic rationale for the use of Bayesian methods is the inherent flexibility introduced by their incorporation of multiple levels of randomness and the resultant ability to combine information from different sources, while incorporating all reasonable sources of uncertainty in inferential summaries. Such methods naturally lead to smoothed estimates in complicated data structures and consequently have the ability to obtain better real-world answers.

Another reason for focusing on Bayesian methods is more psychological, and involves the relationship between the statistician and the client or specialist in the subject matter area who is the consumer of the statistician's work. In many practical cases, clients will interpret interval estimates provided by statisticians as Bayesian intervals, that is, as probability statements about the likely values of unknown quantities conditional on the evidence in the data. Such direct probability statements require prior probability specifications for unknown quantities (or more generally, probability models for vectors of unknowns), and thus the kinds of answers clients will assume are being provided by statisticians, Bayesian answers, require full probability models—explicit or implicit.

Finally, Bayesian inferences are conditional on probability models that invariably contain approximations in their attempt to represent complicated real-world relationships. If the Bayesian answers vary dramatically over a range of scientifically reasonable assumptions that are unassailable by the data, then the resultant range of possible conclusions must be entertained as legitimate, and we believe that the statistician has the responsibility to make the client aware of this fact.

In this book, we focus on the construction of models (especially hierarchical ones, as discussed in Chapter 5 onward) to relate complicated data structures to scientific questions, checking the fit of such models, and investigating the sensitivity of conclusions to reasonable modeling assumptions. From this point of view, the strength of the Bayesian approach lies in (1) its ability to combine information from multiple sources (thereby in fact allowing greater 'objectivity' in final conclusions), and (2) its more encompassing accounting of uncertainty about the unknowns in a statistical problem.

Other important themes, many of which are common to much modern applied statistical practice, whether formally Bayesian or not, are the following:

- a willingness to use many parameters
- hierarchical structuring of models, which is the essential tool for achieving partial pooling of estimates and compromising in a scientific way between alternative sources of information
- model checking—not only by examining the internal goodness of fit of models to observed and possible future data, but also by comparing inferences about estimands and predictions of interest to substantive knowledge
- an emphasis on inference in the form of distributions or at least interval estimates rather than simple point estimates
- the use of simulation as the primary method of computation; the modern computational counterpart to a 'joint probability distribution' is a set of randomly drawn values, and a key tool for dealing with missing data is the method of multiple imputation (computation and multiple imputation are discussed in more detail in later chapters)
- the use of probability models as tools for understanding and possibly improving data-analytic techniques that may not explicitly invoke a Bayesian model
- the importance of including in the analysis as much background information as possible, so as to approximate the goal that data can be viewed as a random sample, conditional on all the variables in the model
- the importance of designing studies to have the property that inferences for estimands of interest will be robust to model assumptions.

1.11 Bibliographic note

Several good introductory books have been written on Bayesian statistics, beginning with Lindley (1965), and continuing through Hoff (2009). Berry (1996) presents, from a Bayesian perspective, many of the standard topics for an introductory statistics textbook. Gill (2002) and Jackman (2009) introduce applied Bayesian statistics for social scientists, Kruschke (2011) introduces Bayesian methods for psychology researchers, and Christensen et al. (2010) supply a general introduction. Carlin and Louis (2008) cover the theory and applications of Bayesian inference, focusing on biological applications and connections to classical methods. Some resources for teaching Bayesian statistics include Sedlmeier and Gigerenzer (2001) and Gelman (1998, 2008b).

The bibliographic notes at the ends of the chapters in this book refer to a variety of specific applications of Bayesian data analysis. Several review articles in the statistical literature, such as Breslow (1990) and Racine et al. (1986), have appeared that discuss, in general terms, areas of application in which Bayesian methods have been useful. The volumes edited by Gatsonis et al. (1993–2002) are collections of Bayesian analyses, including extensive discussions about choices in the modeling process and the relations between the statistical methods and the applications.

The foundations of probability and Bayesian statistics are an important topic that we treat only briefly. Bernardo and Smith (1994) give a thorough review of the foundations of Bayesian models and inference with a comprehensive list of references. Jeffreys (1961) is a self-contained book about Bayesian statistics that comprehensively presents an inductive view of inference; Good (1950) is another important early work. Jaynes (1983) is a collection of reprinted articles that present a deductive view of Bayesian inference that we believe is similar to ours. Both Jeffreys and Jaynes focus on applications in the physical sciences. Jaynes (2003) focuses on connections between statistical inference and the philosophy of science and includes several examples of physical probability.

Gigerenzer and Hoffrage (1995) discuss the connections between Bayesian probability and frequency probabilities from a perspective similar to ours, and provide evidence that people can typically understand and compute best with probabilities that expressed in the form of relative frequency. Gelman (1998) presents some classroom activities for teaching Bayesian ideas.

De Finetti (1974) is an influential work that focuses on the crucial role of exchangeability. More approachable discussions of the role of exchangeability in Bayesian inference are provided by Lindley and Novick (1981) and Rubin (1978a, 1987a). The non-Bayesian article by Draper et al. (1993) makes an interesting attempt to explain how exchangeable probability models can be justified in data analysis. Berger and Wolpert (1984) give a comprehensive discussion and review of the likelihood principle, and Berger (1985, Sections 1.6, 4.1, and 4.12) reviews a range of philosophical issues from the perspective of Bayesian decision theory.

Our own philosophy of Bayesian statistics appears in Gelman (2011) and Gelman and Shalizi (2013); for some contrasting views, see the discussion of that article, along with Efron (1986) and the discussions following Gelman (2008a).

Pratt (1965) and Rubin (1984) discuss the relevance of Bayesian methods for applied statistics and make many connections between Bayesian and non-Bayesian approaches to inference. Further references on the foundations of statistical inference appear in Shafer (1982) and the accompanying discussion. Kahneman and Tversky (1972) and Alpert and Raiffa (1982) present the results of psychological experiments that assess the meaning of 'subjective probability' as measured by people's stated beliefs and observed actions. Lindley (1971a) surveys many different statistical ideas, all from the Bayesian perspective. Box and Tiao (1973) is an early book on applied Bayesian methods. They give an extensive treatment of inference based on normal distributions, and their first chapter, a broad introduction to Bayesian inference, provides a good counterpart to Chapters 1 and 2 of this book.

The iterative process involving modeling, inference, and model checking that we present in Section 1.1 is discussed at length in the first chapter of Box and Tiao (1973) and also in Box (1980). Cox and Snell (1981) provide a more introductory treatment of these ideas from a less model-based perspective.

Many good books on the mathematical aspects of probability theory are available, such as Feller (1968) and Ross (1983); these are useful when constructing probability models and working with them. O'Hagan (1988) has written an interesting introductory text on probability from an explicitly Bayesian point of view.

Physical probability models for coin tossing are discussed by Keller (1986), Jaynes (2003), and Gelman and Nolan (2002b). The football example of Section 1.6 is discussed in more detail in Stern (1991); see also Harville (1980) and Glickman (1993) and Glickman and Stern (1998) for analyses of football scores not using the point spread. Related analyses of sports scores and betting odds appear in Stern (1997, 1998). For more background on sports betting, see Snyder (1975) and Rombola (1984).

An interesting real-world example of probability assignment arose with the explosion of the Challenger space shuttle in 1986; Martz and Zimmer (1992), Dalal, Fowlkes, and Hoadley (1989), and Lavine (1991) present and compare various methods for assigning probabilities for space shuttle failures. (At the time of writing we are not aware of similar contributions relating to the more recent space accident in 2003.) The record-linkage example in Section 1.7 appears in Belin and Rubin (1995b), who discuss the mixture models and calibration techniques in more detail. The Census problem that motivated the record linkage is described by Hogan (1992).

In all our examples, probabilities are assigned using statistical modeling and estimation, not by 'subjective' assessment. Dawid (1986) provides a general discussion of probability assignment, and Dawid (1982) discusses the connections between calibration and Bayesian probability assignment.

The graphical method of jittering, used in Figures 1.1 and 1.2 and elsewhere in this book, is discussed in Chambers et al. (1983). For information on the statistical packages R and Bugs, see Becker, Chambers, and Wilks (1988), R Project (2002), Fox (2002), Venables and Ripley (2002), and Spiegelhalter et al. (1994, 2003).

Norvig (2007) describes the principles and details of the Bayesian spelling corrector.

1.12 Exercises

1. Conditional probability: suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $\Pr(\theta = 1) = 0.5$ and $\Pr(\theta = 2) = 0.5$.
   (a) For $\sigma = 2$, write the formula for the marginal probability density for y and sketch it.
   (b) What is $\Pr(\theta = 1 | y = 1)$, again supposing $\sigma = 2$?
   (c) Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.
2. Conditional means and variances: show that (1.8) and (1.9) hold if u is a vector.
3. Probability calculation for genetics (from Lindley, 1965): suppose that in each individual of a large population there is a pair of genes, each of which can be either x or X, that controls eye color: those with xx have blue eyes, while heterozygotes (those with Xx or xX) and those with XX have brown eyes. The proportion of blue-eyed individuals is p^2 and of heterozygotes is $2p(1 - p)$, where $0 < p < 1$. Each parent transmits one of its own genes to the child; if a parent is a heterozygote, the probability that it transmits the gene of type X is $\frac{1}{2}$. Assuming random mating, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is $2p/(1 + 2p)$. Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children, all brown-eyed. Find the posterior probability that Judy is a heterozygote and the probability that her first grandchild has blue eyes.
4. Probability assignment: we will use the football dataset to estimate some conditional probabilities about professional football games. There were twelve games with point spreads of 8 points; the outcomes in those games were: -7, -5, -3, -3, 1, 6, 7, 13, 15, 16, 20, and 21, with positive values indicating wins by the favorite and negative values indicating wins by the underdog. Consider the following conditional probabilities:

$$\Pr(\text{favorite wins} \,|\, \text{point spread} = 8),$$
$$\Pr(\text{favorite wins by at least } 8 \,|\, \text{point spread} = 8),$$
$$\Pr(\text{favorite wins by at least } 8 \,|\, \text{point spread} = 8 \text{ and favorite wins}).$$

   (a) Estimate each of these using the relative frequencies of games with a point spread of 8.
   (b) Estimate each using the normal approximation for the distribution of (outcome $-$ point spread).
5. Probability assignment: the 435 U.S. Congressmembers are elected to two-year terms; the number of voters in an individual congressional election varies from about 50,000 to 350,000. We will use various sources of information to estimate roughly the probability that at least one congressional election is tied in the next national election.
   (a) Use any knowledge you have about U.S. politics. Specify clearly what information you are using to construct this conditional probability, even if your answer is just a guess.
   (b) Use the following information: in the period 1900–1992, there were 20,597 congressional elections, out of which 6 were decided by fewer than 10 votes and 49 decided by fewer than 100 votes.

See Gelman, King, and Boscardin (1998), Mulligan and Hunter (2001), and Gelman, Katz, and Tuerlinckx (2002) for more on this topic.

6. Conditional probability: approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as $\frac{1}{2}$.)
7. Conditional probability: the following problem is loosely based on the television game show *Let's Make a Deal*. At the end of the show, a contestant is asked to choose one of three large boxes, where one box contains a fabulous prize and the other two boxes contain lesser prizes. After the contestant chooses a box, Monty Hall, the host of the show, opens one of the two boxes containing smaller prizes. (In order to keep the conclusion suspenseful, Monty does not open the box selected by the contestant.) Monty offers the contestant the opportunity to switch from the chosen box to the remaining unopened box. Should the contestant switch or stay with the original choice? Calculate the probability that the contestant wins under each strategy. This is an exercise in being clear about the information that should be conditioned on when constructing a probability judgment. See Selvin (1975) and Morgan et al. (1991) for further discussion of this problem.
8. Subjective probability: discuss the following statement. 'The probability of event E is considered "subjective" if two rational persons A and B can assign unequal probabilities to E, $P_A(E)$ and $P_B(E)$. These probabilities can also be interpreted as "conditional": $P_A(E) = P(E|I_A)$ and $P_B(E) = P(E|I_B)$, where I_A and I_B represent the knowledge available to persons A and B, respectively.' Apply this idea to the following examples.
   (a) The probability that a '6' appears when a fair die is rolled, where A observes the outcome of the die roll and B does not.
   (b) The probability that Brazil wins the next World Cup, where A is ignorant of soccer and B is a knowledgeable sports fan.
9. Simulation of a queuing problem: a clinic has three doctors. Patients come into the clinic at random, starting at 9 a.m., according to a Poisson process with time parameter 10 minutes: that is, the time after opening at which the first patient appears follows an exponential distribution with expectation 10 minutes and then, after each patient arrives, the waiting time until the next patient is independently exponentially distributed, also with expectation 10 minutes. When a patient arrives, he or she waits until a doctor is available. The amount of time spent by each doctor with each patient is a random variable, uniformly distributed between 5 and 20 minutes. The office stops admitting new patients at 4 p.m. and closes when the last patient is through with the doctor.
   (a) Simulate this process once. How many patients came to the office? How many had to wait for a doctor? What was their average wait? When did the office close?
   (b) Simulate the process 100 times and estimate the median and 50% interval for each of the summaries in (a).

References

The literature of Bayesian statistics is vast, especially in recent years. Instead of trying to be exhaustive, we supply here a selective list of references that may be useful for applied Bayesian statistics. Many of these sources have extensive reference lists of their own, which may be useful for an in-depth exploration of a topic. We also include references from non-Bayesian statistics and numerical analysis that present probability models or calculations relevant to Bayesian methods.

Abayomi, K., Gelman, A., and Levy, M. (2008). Diagnostics for multivariate imputations. *Applied Statistics* **57**, 273–291.

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). The Gaussian process density sampler. In *Advances in Neural Information Processing Systems 21*, ed. D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, 9–16.

Agresti, A. (2002). *Categorical Data Analysis*, second edition. New York: Wiley.

Agresti, A., and Coull, B. A. (1998). Approximate is better than exact for interval estimation of binomial proportions. *American Statistician* **52**, 119–126.

Aitchison, J., and Dunsmore, I. R. (1975). *Statistical Prediction Analysis.* Cambridge University Press.

Aitkin, M., and Longford, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society A* **149**, 1–43.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, ed. B. N. Petrov and F. Csaki, 267–281. Budapest: Akademiai Kiado. Reprinted in *Breakthroughs in Statistics*, ed. S. Kotz, 610–624. New York: Springer (1992).

Albert, J. H. (1988). Bayesian estimation of Poisson means using a hierarchical log-linear model. In *Bayesian Statistics 3*, ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, 519–531. Oxford University Press.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics* **17**, 251–269.

Albert, J. H., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Albert, J. H., and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–759.

Alpert, M., and Raiffa, H. (1982). A progress report on the training of probability assessors. In *Judgment Under Uncertainty: Heuristics and Biases*, ed. Kahneman, D., Slovic, P., and Tversky, A., 294–305. Cambridge University Press.

Anderson, D. A. (1988). Some models for overdispersed binomial data. *Australian Journal of Statistics* **30**, 125–148.

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association* **52**, 200–203.

Ando, T., and Tsay, R. (2010). Predictive likelihood for Bayesian model selection and averaging. *International Journal of Forecasting* **26**, 744–763.

Andrieu, C., and Robert, C. (2001). Controlled MCMC for optimal sampling. Technical report, Department of Mathematics, University of Bristol.

Andrieu, C., and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing* **18**, 343–373.

Angrist, J., Imbens, G., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

Anscombe, F. J. (1963). Sequential medical trials. *Journal of the American Statistical Association* **58**, 365–383.

Ansolabehere, S., and Snyder, J. M. (2002). The incumbency advantage in U.S. elections: An analysis of state and federal offices, 1942–2000. *Election Law Journal* **1**, 315–338.

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys* **4**, 40–79.

Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica* **23**, 119–143.

Armagan, A., Dunson, D. B., Lee, J., and Bajwa, W. U. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*.

Arminger, G. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271–300.

Atkinson, A. C. (1985). *Plots, Transformations, and Regression.* Oxford University Press.

Banerjee, A., Dunson, D. B., and Tokdar, S. (2011). Efficient Gaussian process regression for large data sets. `http://arxiv.org/abs/1106.5779`

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). Hierarchical modeling and analysis for spatial data. London: Chapman & Hall.

Barbieri, M. M., and Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics* **32**, 870–897.

Barnard, G. A. (1949). Statistical inference (with discussion). *Journal of the Royal Statistical Society B* **11**, 115–139.

Barnard, G. A. (1985). Pivotal inference. In *Encyclopedia of Statistical Sciences*, Vol. 6, ed. S. Kotz, N. L. Johnson, and C. B. Read, 743–747. New York: Wiley.

Barnard, J., Frangakis, C., Hill, J., and Rubin, D. B. (2003). A principal stratification approach to broken randomized experiments: A case study of vouchers in New York City (with discussion). *Journal of the American Statistical Association.*

Barnard, J., McCulloch, R. E., and Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10**, 1281–1311.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of ring-recovery data using random effects. *Biometrics* **59**, 54–65.

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications.* New York: Wiley.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* **41**, 164–171.

Bayarri, M. J., and Berger, J. (1998). Quantifying surprise in the data and model verification (with discussion). In *Bayesian Statistics 6*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 53–82. Oxford University Press.

Bayarri, M. J., and Berger, J. (2000). P-values for composite null models (with discussion). *Journal of the American Statistical Association* **95**, 1127–1142.

Bayarri, M. J., and Castellanos, M. E. (2007). Bayesian checking of the second levels of hierarchical models (with discussion). *Statistical Science* **22**, 322–367.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 330–418. Reprinted, with biographical note by G. A. Barnard, in *Biometrika* **45**, 293–315 (1958).

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics.* Pacific Grove, Calif.: Wadsworth.

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association* **91**, 1450–1460.

Belin, T. R., Diffendal, G. J., Mack, S., Rubin, D. B., Schafer, J. L., and Zaslavsky, A. M.

(1993). Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation (with discussion). *Journal of the American Statistical Association* **88**, 1149–1166.

Belin, T. R., and Rubin, D. B. (1990). Analysis of a finite mixture model with variance components. In *Proceedings of the American Statistical Association, Social Statistics Section*, 211–215.

Belin, T. R., and Rubin, D. B. (1995a). The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statistics in Medicine* **14**, 747–768.

Belin, T. R., and Rubin, D. B. (1995b). A method for calibrating false-match rates in record linkage. *Journal of the American Statistical Association* **90**, 694–707.

Berger, J. O. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness in Bayesian Statistics*, ed. J. Kadane. Amsterdam: North-Holland.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition. New York: Springer.

Berger, J. O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference* **25**, 303–328.

Berger, J. O., and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with epsilon-contaminated priors. *Annals of Statistics* **14**, 461–486.

Berger, J. O., and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of P values and evidence (with discussion). *Journal of the American Statistical Association* **82**, 112–139.

Berger, J. O., and Wolpert, R. (1984). *The Likelihood Principle*. Hayward, Calif.: Institute of Mathematical Statistics.

Berkhof, J., Van Mechelen, I., and Gelman, A. (2003). A Bayesian approach to the selection and testing of latent class models. *Statistica Sinica* **13**, 423–442.

Bernardinelli, L., Clayton, D. G., and Montomoli, C. (1995). Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine* **14**, 2411–2431.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society B* **41**, 113–147.

Bernardo, J. M., and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.

Berry, D. A. (1996). *Statistics: A Bayesian Perspective*. Belmont, Calif.: Wadsworth.

Berry, S., M., Carlin, B. P., Lee, J. J., and Muller, P. (2010). *Bayesian Adaptive Methods for Clinical Trials*. London: Chapman & Hall.

Berzuini, C., Best, N. G., Gilks, W. R., and Larizza, C. (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association* **92**, 1403–1412.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B* **36**, 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society B* **48**, 259–302.

Besag, J., and Green, P. J. (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society B* **55**, 25–102.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

Besag, J., and Higdon, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society B* **61**, 691–746.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Betancourt, M. J. (2013). A general metric for Riemannian manifold Hamiltonian Monte Carlo. In *Geometric science of information*, ed. F. Nielsen, and F. Barbaresco, 327–334.

Betancourt, M. J. (2013). Generalizing the no-U-turn sampler to Riemannian manifolds. `http://arxiv.org/abs/1304.1920`

Betancourt, M. J., and Stein, L. C. (2011). The geometry of Hamiltonian Monte Carlo. `http://arxiv.org/abs/1112.4118`

Bickel, P., and Blackwell, D. (1967). A note on Bayes estimates. *Annals of Mathematical Statistics* **38**, 1907–1911.

Bigelow, J. L., and Dunson, D. B. (2009). Bayesian semiparametric joint models for functional predictors. *Journal of the American Statistical Association* **104**, 26–36.

Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics* **9**, 122–140.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Blei, D., Ng, A., and Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research* **3**, 993–1022.

Bloom, H. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review* **8**, 225–246.

Bock, R. D., ed. (1989). *Multilevel Analysis of Educational Data*. New York: Academic Press.

Boscardin, W. J., and Gelman, A. (1996). Bayesian regression with parametric models for heteroscedasticity. *Advances in Econometrics* **11A**, 87–109.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A* **143**, 383–430.

Box, G. E. P. (1983). An apology for ecumenism in statistics. In *Scientific Inference, Data Analysis, and Robustness*, ed. G. E. P. Box, T. Leonard, T., and C. F. Wu, 51–84. New York: Academic Press.

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B* **26**, 211–252.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, second edition. San Francisco: Holden-Day.

Box, G. E. P., and Tiao, G. C. (1962). A further look at robustness via Bayes's theorem. *Biometrika* **49**, 419–432.

Box, G. E. P., and Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika* **55**, 119–129.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics.

Bradley, R. A., and Terry, M. E. (1952). The rank analysis of incomplete block designs. 1. The method of paired comparisons. *Biometrika* **39**, 324–345.

Bradlow, E. T., and Fader, P. S. (2001). A Bayesian lifetime model for the "Hot 100" Billboard songs. *Journal of the American Statistical Association* **96**, 368–381.

Braun, H. I., Jones, D. H., Rubin, D. B., and Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. *Psychometrika* **48**, 171–181.

Breslow, N. (1990). Biostatistics and Bayes (with discussion). *Statistical Science* **5**, 269–298.

Bretthorst, G. L. (1988). *Bayesian Spectrum Analysis and Parameter Estimation*. New York: Springer.

Brewer, K. W. R. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics* **5**, 93–105.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, expanded edition. San Francisco: Holden-Day.

Brooks, S. P., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

Brooks, S. P., and Giudici, P. (2000). MCMC convergence assessment via two-way ANOVA. *Journal of Computational and Graphical Statistics* **9**, 266–285.

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society B* **65**, 3–55.

Brooks, S. P., and Roberts, G. O. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing* **8**, 319–335.

Browner, W. S., and Newman, T. B. (1987). Are all significant P values created equal? *Journal of the American Medical Association* **257**, 2459–2463.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* **76**, 503–514.

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach.* New York: Springer.

Bush, R. R., and Mosteller, F. (1955). *Stochastic Models for Learning.* New York: Wiley.

Calvin, J. A., and Sedransk, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *Journal of the American Statistical Association* **86**, 36–48.

Carlin, B. P., and Chib, S. (1993). Bayesian model choice via Markov chain Monte Carlo. *Journal of the Royal Statistical Society B* **57**, 473–484.

Carlin, B. P., and Gelfand, A. E. (1993). Parametric likelihood inference for record breaking problems. *Biometrika* **80**, 507–515.

Carlin, B. P., and Louis, T. A. (2008). *Bayesian Methods for Data Analysis*, third edition. New York: Chapman & Hall.

Carlin, B. P., and Polson, N. G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler. *Canadian Journal of Statistics* **19**, 399–405.

Carlin, J. B. (1992). Meta-analysis for 2×2 tables: A Bayesian approach. *Statistics in Medicine* **11**, 141–158.

Carlin, J. B., and Dempster, A. P. (1989). Sensitivity analysis of seasonal adjustments: empirical case studies (with discussion). *Journal of the American Statistical Association* **84**, 6–32.

Carlin, J. B., Stevenson, M. R., Roberts, I., Bennett, C. M., Gelman, A., and Nolan, T. (1997). Walking to school and traffic exposure in Australian children. *Australian and New Zealand Journal of Public Health* **21**, 286–292.

Carlin, J. B., Wolfe, R., Brown, C. H., and Gelman, A. (2001). A case study on the choice, interpretation, and checking of multilevel models for longitudinal binary outcomes. *Biostatistics* **2**, 397–416.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* New York: Chapman & Hall.

Carvalho, C. M., Lopes, H. F., Polson, N. G., and Taddy, M. A. (2010). Particle learning for general mixtures. *Bayesian Analysis* **5**, 709–740.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480.

Chaloner, K. (1991). Bayesian residual analysis in the presence of censoring. *Biometrika* **78**, 637–644.

Chaloner, K., and Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika* **75**, 651–659.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis.* Pacific Grove, Calif.: Wadsworth.

Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation.* New York: Springer.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design.* Philadelphia: Society for Industrial and Applied Mathematics.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician* **49**, 327–335.

Chib, S., and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96**, 270–281.

Chipman, H., George, E. I., and McCulloch, R. E. (1998). Bayesian CART model search (with discussion). *Journal of the American Statistical Association* **93**, 935–960.

Chipman, H., George, E. I., and McCulloch, R. E. (2001). The practical implementation of Bayesian model selection (with discussion). In *Model Selection* (Institute of Mathematical Statistics Lecture Notes 38), ed. P. Lahiri, 67–116.

Chipman, H., George, E. I., and McCulloch, R. E. (2002). Bayesian treed models. *Machine Learning* **48**, 299–320.

Chipman, H., Kolaczyk, E., and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association* **92**, 1413–1421.

Christensen, R., Johnson, W. O., Branscum, A. J., and Hanson, T. E. (2010). *Bayesian Ideas and Data Analysis*. London: Chapman & Hall.

Chung, Y., and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* **104**, 1646–1660.

Chung, Y., and Dunson, D. B. (2011). The local Dirichlet process. *Annals of the Institute of Statistical Mathematics* **63**, 59–80.

Chung, Y., Rabe-Hesketh, S., Gelman, A., Liu, J. C., and Dorie, A. (2013a). A non-degenerate penalized likelihood estimator for hierarchical variance parameters in multilevel models. *Psychometrika*.

Chung, Y., Rabe-Hesketh, S., Gelman, A., Liu, J. C., and Dorie, A. (2013b). Nonsingular covariance estimation in linear mixed models through weakly informative priors. Technical report, School of Education, University of California, Berkeley.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467–485.

Clayton, D. G., and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, ed. P. Elliott, J. Cusick, D. English, and R. Stern, 205–220. Oxford University Press.

Clayton, D. G., and Kaldor, J. M. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–682.

Clemen, R. T. (1996). *Making Hard Decisions*, second edition. Belmont, Calif.: Duxbury Press.

Cleveland, W. S. (1985). *The Elements of Graphing Data*. Monterey, Calif.: Wadsworth.

Cleveland, W. S. (1993). *Envisioning Information*. Summit, N.J.: Hobart.

Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B., and Wideman, L. (1991). Multiple imputation of industry and occupation codes in Census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* **86**, 68–78.

Clyde, M., DeSimone, H., and Parmigiani, G. (1996). Prediction via orthogonalized model mixing. *Journal of the American Statistical Association* **91**, 1197–1208.

Connors, A. F., Speroff, T., Dawson, N. V., Thomas, C., Harrell, F. E., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., Fulkerson, W. J., Vidaillet, H., Broste, S., Bellamy, P., Lynn, J., and Knauss, W. A. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. *Journal of the American Medical Association* **276**, 889–997.

Conover, W. J., and Iman, R. L. (1980). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician* **35**, 124–129.

Cook, S., Gelman, A., and Rubin, D. B. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* **15**, 675–692.

Cowles, M. K., and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 833–904.

Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics*. New York: Chapman & Hall.

Cox, D. R., and Snell, E. J. (1981). *Applied Statistics*. New York: Chapman & Hall.

Cox, G. W., and Katz, J. (1996). Why did the incumbency advantage grow? *American Journal of Political Science* **40**, 478–497.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, second edition. New York: Wiley.

Cressie, N. A. C., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications* **19**, 553–570.

Cseke, B., and Heskes, T. (2011). Approximate marginals in latent Gaussian models. *Journal of Machine Learning Research* **12**, 417–454.

Dalal, S. R., Fowlkes, E. B., and Hoadley, B. (1989). Risk analysis of the space shuttle: pre-Challenger prediction of failure. *Journal of the American Statistical Association* **84**, 945–957.

Daniels, M. J., and Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association* **94**, 1254–1263.

Daniels, M. J., and Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57**, 1173–1184.

Daniels, M. J., and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* **89**, 553–566.

Datta, G. S., Lahiri, P., Maiti, T., and Lu, K. L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association* **94**, 1074–1082.

Daume, H. (2008). HBC: Hierarchical Bayes compiler.
`http://www.umiacs.umd.edu/~hal/HBC/hbc.pdf`

David, H. A. (1988). *The Method of Paired Comparisons*, second edition. Oxford University Press.

David, M. H., Little, R. J. A., Samuhel, M. E., and Triest, R. K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association* **81**, 29–41.

Davidson, R. R., and Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics* **33**, 693–702.

Dawid, A. P. (1982). The well-calibrated Bayesian (with discussion). *Journal of the American Statistical Association* **77**, 605–610.

Dawid, A. P. (1986). Probability forecasting. In *Encyclopedia of Statistical Sciences*, Vol. 7, ed. S. Kotz, N. L. Johnson, and C. B. Read, 210–218. New York: Wiley.

Dawid, A. P. (2000). Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association* **95**, 407–448.

Dawid, A. P., and Dickey, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *Journal of the American Statistical Association* **72**, 845–850.

Dawid, A. P., Stone, M., and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inferences (with discussion). *Journal of the Royal Statistical Society B* **35**, 189–233.

Deely, J. J., and Lindley, D. V. (1981). Bayes empirical Bayes. *Journal of the American Statistical Association* **76**, 833–841.

de Finetti, B. (1974). *Theory of Probability.* New York: Wiley.

DeGroot, M. H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

Dehejia, R. (2005). Program evaluation as a decision problem. *Journal of Econometrics* **125**, 141–173.

Dehejia, R., and Wahba, S. (1999). Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**, 1053–1062.

De Iorio, M., Muller, P., Rosner, G. L., and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

De la Cruz-Mesia, R., Quintana, F., and Muller, P. (2009). Semiparametric Bayesian classification with longitudinal markers. *Applied Statistics* **56**, 119–137.

Dellaportas, P., and Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics* **42**, 443–459.

Deming, W. E., and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* **11**, 427–444.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **38**, 205–247.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society B* **30**, 205–247.

Dempster, A. P. (1971). Model searching and estimation in the logic of inference. In *Proceedings of the Symposium on the Foundations of Statistical Inference*, ed. V. P. Godambe and D. A. Sprott, 56–81. Toronto: Holt, Rinehart and Winston.

Dempster, A. P. (1975). A subjectivist look at robustness. *Bulletin of the International Statistical Institute* **46**, 349–374.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**, 1–38.

Dempster, A. P., and Raghunathan, T. E. (1987). Using a covariate for small area estimation: A common sense Bayesian approach. In *Small Area Statistics: An International Symposium*, ed. R. Platek, J. N. K. Rao, C. E. Sarndal, and M. P. Singh, 77–90. New York: Wiley.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association* **76**, 341–353.

Dempster, A. P., Selwyn, M. R., and Weeks, B. J. (1983). Combining historical and randomized controls for assessing trends in proportions. *Journal of the American Statistical Association* **78**, 221–227.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., and Smith, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression.* New York: Wiley.

Diebolt, J., and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B* **56**, 363–375.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055–1071.

Dobra, A., Tebaldi, C., and West, M. (2003). Bayesian inference for incomplete multi-way tables. Technical report, Institute of Statistics and Decision Sciences, Duke University.

Dominici, F., Parmigiani, G., Wolpert, R. L., and Hasselblad, V. (1999). Meta-analysis of migraine headache treatments: combining information from heterogeneous designs. *Journal of the American Statistical Association* **94**, 16–28.

Donoho, D. L., Johnstone, I. M., Hoch, J. C., and Stern, A. S. (1992). Maximum entropy and the nearly black object (with discussion). *Journal of the Royal Statistical Society B* **54**, 41–81.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society B* **57**, 45–97.

Draper, D., Hodges, J. S., Mallows, C. L., and Pregibon, D. (1993). Exchangeability and data analysis. *Journal of the Royal Statistical Society A* **156**, 9–37.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195**, 216–222.

DuMouchel, W. M. (1990). Bayesian meta-analysis. In *Statistical Methodology in the Pharmaceutical Sciences*, ed. D. A. Berry, 509–529. New York: Marcel Dekker.

DuMouchel, W. M., and Harris, J. E. (1983). Bayes methods for combining the results of cancer studies in humans and other species (with discussion). *Journal of the American Statistical Association* **78**, 293–315.

Dunson, D. B. (2005). Bayesian semiparametric isotonic regression for count data. *Journal of the American Statistical Association* **100**, 618–627.

Dunson, D. B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics* **7**, 551–568.

Dunson, D. B. (2009). Bayesian nonparametric hierarchical modeling. *Biometrical Journal* **51**, 273–284.

Dunson, D. B. (2010a). Flexible Bayes regression of epidemiologic data. In *Oxford Handbook of Applied Bayesian Analysis*, ed. A. O'Hagan and M. West. Oxford University Press.

Dunson, D. B. (2010b). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics*, ed. N. L. Hjort, C. Holmes, P. Muller, and S. G. Walker. Cambridge University Press.

Dunson, D. B., and Bhattacharya, A. (2010). Nonparametric Bayes regression and classification through mixtures of product kernels. In *Bayesian Statistics 9*, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, 145–164. Oxford University Press.

Dunson, D. B., Pillai, N., and Park, J. H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society B* **69**, 163–183.

Dunson, D. B., and Park, J. H. (2009). Kernel stick-breaking processes. *Biometrika* **95**, 307–323.

Dunson, D. B., and Peddada, S. D. (2008). Bayesian nonparametric inference on stochastic ordering. *Biometrika* **95**, 859–874.

Dunson, D. B., and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics* **17**, 385–400.

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review* **70**, 193–242.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403–417.

Efron, B. (1986). Why isn't everyone a Bayesian? *American Statistician* **40**, 1–5.

Efron, B., and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators—Part I: The Bayes case. *Journal of the American Statistical Association* **66**, 807–815.

Efron, B., and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators—Part II: The empirical Bayes case. *Journal of the American Statistical Association* **67**, 130–139.

Efron, B., and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* **70**, 311–319.

Efron, B., and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63**, 435–448.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Efron, B., and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology* **23**, 70–86.

Ehrenberg, A. S. C. (1986). Discussion of Racine et al. (1986). *Applied Statistics* **35**, 135–136.

Ericson, W. A. (1969). Subjective Bayesian models in sampling finite populations, I. *Journal of the Royal Statistical Society B* **31**, 195–234.

Fay, R. E., and Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74**, 269–277.

Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62**, 89–100.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 1, third edition. New York: Wiley.

Fienberg, S. E. (1977). *The Analysis of Cross-Classified Categorical Data*. Cambridge, Mass.: MIT Press.

Fienberg, S. E. (2000). Contingency tables and log-linear models: basic results and new developments. *Journal of the American Statistical Association* **95**, 643–647.

Fill, J. A. (1998). An interruptible algorithm for perfect sampling. *Annals of Applied Probability* **8**, 131–162.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society* **222**, 309–368.

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**, 42–58.

Ford, E. S., Kelly, A. E., Teutsch, S. M., Thacker, S. B., and Garbe, P. L. (1999). Radon and lung cancer: A cost-effectiveness analysis. *American Journal of Public Health* **89**, 351–357.

Fouskakis, D., and Draper, D. (2008). Comparing stochastic optimization methods for variable selection in binary outcome prediction with application to health policy. *Journal of the American Statistical Association* **103**, 1367–1381.

Fouskakis, D., Ntzoufras, I., and Draper, D. (2009). Population-based reversible-jump Markov chain Monte Carlo for Bayesian variable selection and evaluation under cost limit restrictions. *Applied Statistics* **58**, 383–403.

Fox, J. (2002). *An R and S-Plus Companion to Applied Regression*. London: Sage.

Fraley, C., and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611–631.

Frangakis, C., and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Freedman, L. S., Spiegelhalter, D. J., and Parmar, M. K. B. (1994). The what, why and how of Bayesian clinical trials monitoring. *Statistics in Medicine* **13**, 1371–1383.

Gatsonis, C., Hodges, J. S., Kass, R. E., Singpurwalla, N. D., West, M., Carlin, B. P., Carriquiry, A., Gelman, A., Pauler, D., Verdinelli, I., and Wakefield, J., eds. (1993–2002). *Case Studies in Bayesian Statistics*, volumes 1–7. New York: Springer.

Gaver, D. P., and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analyses of event rates. *Technometrics* **29**, 1–15.

Geisser, S. (1986). Predictive analysis. In *Encyclopedia of Statistical Sciences*, Vol. 7, ed. S. Kotz, N. L. Johnson, and C. B. Read, 158–170. New York: Wiley.

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, D. J. Spiegelhalter, 145–162. London: Chapman & Hall.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 147–167. Oxford University Press.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

Gelfand, A. E., and Sahu, S. K. (1994). On Markov chain Monte Carlo acceleration. *Journal of Computational and Graphical Statistics* **3**, 261–276.

Gelfand, A. E., and Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247–253.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parameterizations for normal linear mixed models. *Biometrika* **82**, 479–488.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A. (1992a). Discussion of 'Maximum entropy and the nearly black object,' by Donoho et al. *Journal of the Royal Statistical Society B* **54**, 72.

Gelman, A. (1992b). Iterative and non-iterative simulation algorithms. *Computing Science and Statistics* **24**, 433–438.

Gelman, A. (1998). Some class-participation demonstrations for decision theory and Bayesian statistics. *American Statistician* **52**, 167–174.

Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review* **71**, 369–382.

Gelman, A. (2004a). Exploratory data analysis for complex models (with discussion). *Journal of Computational and Graphical Statistics* **13**, 755–787.

Gelman, A. (2004b). Parameterization and Bayesian modeling. *Journal of the American Statistical Association* **99**, 537–545.

Gelman, A. (2005). Analysis of variance: why it is more important than ever (with discussion). *Annals of Statistics* **33**, 1–53.

Gelman, A. (2006a). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

Gelman, A. (2006b). The boxer, the wrestler, and the coin flip: A paradox of robust Bayesian inference and belief functions. *American Statistician* **60**, 146–150.

Gelman, A. (2007a). Struggles with survey weighting and regression modeling (with discussion). *Statistical Science* **22**, 153–188.

Gelman, A. (2007b). Discussion of 'Bayesian checking of the second levels of hierarchical models,' by M. J. Bayarri and M. E. Castellanos. *Statistical Science* **22**, 349–352.

Gelman, A. (2008a). Objections to Bayesian statistics (with discussion). *Bayesian Analysis* **3**, 445–478.

Gelman, A. (2008b). Teaching Bayesian applied statistics to graduate students in political science, sociology, public health, education, economics, ... *American Statistician* **62**, 202–205.

Gelman, A. (2011). Induction and deduction in Bayesian data analysis. *Rationality, Markets and Morals*, special topic issue 'Statistical science and philosophy of science: Where do (should) they meet in 2011 and beyond?', ed. D. Mayo, A. Spanos, and K. Staley.

Gelman, A. (2013a). P-values and statistical practice. *Epidemiology* **24**, 69–72.

Gelman, A. (2013b). Two simple examples for understanding posterior p-values whose distributions are far from unform. *Electronic Journal of Statistics* **7**, 2595–2602.

Gelman, A., Bois, F. Y., and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* **91**, 1400–1412.

Gelman, A., and Carlin, J. B. (2001). Poststratification and weighting adjustments. In *Survey Nonresponse*, ed. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little. New York: Wiley.

Gelman, A., Chew, G. L., and Shnaidman, M. (2004). Bayesian analysis of serial dilution assays. *Biometrics* **60**, 407–417.

Gelman, A., Fagan, J., and Kiss, A. (2007). An analysis of the NYPD's stop-and-frisk policy in the context of claims of racial bias. *Journal of the American Statistical Association* **102**, 813–823.

Gelman, A., Goegebeur, Y., Tuerlinckx, F., and Van Mechelen, I. (2000). Diagnostic checks for discrete-data regression models using posterior predictive simulations. *Applied Statistics* **49**, 247–268.

Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., Hill, J., and Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* **5**, 189–211.

Gelman, A., and Huang, Z. (2008). Estimating incumbency advantage and its variation, as an example of a before-after study (with discussion). *Journal of the American Statistical Association* **103**, 437–451.

Gelman, A., Huang, Z., van Dyk, D. A., and Boscardin, W. J. (2008). Using redundant parameters to fit hierarchical models. *Journal of Computational and Graphical Statistics* **17**, 95–122.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24**, 997–1016.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**, 1360–1383.

Gelman, A., Katz, J. N., and Tuerlinckx, F. (2002). The mathematics and statistics of voting power. *Statistical Science* **17**, 420–435.

Gelman, A., and King, G. (1990a). Estimating incumbency advantage without bias. *American Journal of Political Science* **34**, 1142–1164.

Gelman, A., and King, G. (1990b). Estimating the electoral consequences of legislative redistricting. *Journal of the American Statistical Association* **85**, 274–282.

Gelman, A., and King, G. (1993). Why are American Presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science* **23**, 409–451.

Gelman, A., King, G., and Boscardin, W. J. (1998). Estimating the probability of events that have never occurred: when does your vote matter? *Journal of the American Statistical Association* **93**, 1–9.

Gelman, A., King, G., and Liu, C. (1998). Multiple imputation for multiple surveys (with discussion). *Journal of the American Statistical Association* **93**, 846–874.

Gelman, A., and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* **23**, 127–135.

Gelman, A., and Meng, X. L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185.

Gelman, A., Meng, X. L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6**, 733–807.

Gelman, A., and Nolan, D. (2002a). *Teaching Statistics: A Bag of Tricks*. Oxford University Press.

Gelman, A., and Nolan, D. (2002b). You can load a die but you can't bias a coin. *American Statistician* **56**, 308–311.

Gelman, A., and Nolan, D. (2002c). A probability model for golf putting. *Teaching Statistics* **24**, 93–95.

Gelman, A., and Price, P. N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine* **18**, 3221–3234.

Gelman, A., and Raghunathan, T. E. (2001). Using conditional distributions for missing-data imputation. Discussion of 'Conditionally specified distributions,' by Arnold et al. *Statistical Science* **3**, 268–269.

Gelman, A., Roberts, G., and Gilks, W. (1995). Efficient Metropolis jumping rules. In *Bayesian Statistics 5*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 599–607. Oxford University Press.

Gelman, A., and Rubin, D. B. (1991). Simulating the posterior distribution of loglinear contingency table models. Technical report.

Gelman, A., and Rubin, D. B. (1992a). A single sequence from the Gibbs sampler gives a false sense of security. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 625–631. Oxford University Press.

Gelman, A., and Rubin, D. B. (1992b). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.

Gelman, A., and Rubin, D. B. (1995). Avoiding model selection in Bayesian social research. Discussion of Raftery (1995b). In *Sociological Methodology 1995*, ed. P. V. Marsden, 165–173.

Gelman, A., and Shalizi, C. (2013). Philosophy and the practice of Bayesian statistics (with discussion). *British Journal of Mathematical and Statistical Psychology* **66**, 8–80.

Gelman, A., and Shirley, K. (2011). Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo*, ed. S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, 163–174. New York: Chapman & Hall.

Gelman, A., Shor, B., Bafumi, J., and Park, D. K. (2007). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Quarterly Journal of Political Science* **2**, 345–367.

Gelman, A., Stevens, M., and Chan, V. (2003). Regression modeling and meta-analysis for decision making: A cost-benefit analysis of a incentives in telephone surveys. *Journal of Business and Economic Statistics* **21**, 213–225.

Gelman, A., and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics* **15**, 373–390.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics* **61**, 74–85.

Gelman, A., and Weakliem, D. (2009). Of beauty, sex, and power: Statistical challenges in estimating small effects. *American Scientist* **97**, 310–316.

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Genovese, C. R. (2001). A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). *Journal of the American Statistical Association* **95**, 691–703.

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods*, second edition. New York: Springer.

George, E. I., and Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87**, 731–747.

George, E. I., and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

George, E. I., and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

Gershman, S. J., Hoffman, M. D., and Blei, D. M. (2012). Nonparametric variational inference. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**, 1317–1339.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. *Computing Science and Statistics* **23**, 156–163.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–483.

Geyer, C. J., and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society B* **54**, 657–699.

Geyer, C. J., and Thompson, E. A. (1993). Annealing Markov chain Monte Carlo with applications to pedigree analysis. Technical report, School of Statistics, University of Minnesota.

Ghitza, Y., and Gelman, A. (2013). Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *American Journal of Political Science* **57**, 762–776.

Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* **102**, 684–704.

Gilks, W. R., Best, N., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* **44**, 455–472.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J. (1993). Modelling complexity: Applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society B* **55**, 39–102.

Gilks, W. R., Richardson, S., and Spiegelhalter, D., eds. (1996). *Practical Markov Chain Monte Carlo*. New York: Chapman & Hall.

Gilks, W. R., and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

Gill, J. (2002). *Bayesian Methods for the Social and Behavioral Sciences*. New York: Chapman & Hall.

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical Optimization*. New York: Academic Press.

Gilovich, T., Griffin, D., and Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

Giltinan, D., and Davidian, M. (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman & Hall.

Girolami, M., and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B* **73**, 123–214.

Glickman, M. E. (1993). Paired comparison models with time-varying parameters. Ph.D. thesis, Department of Statistics, Harvard University.

Glickman, M. E., and Normand, S. L. (2000). The derivation of a latent threshold instrumental variables model. *Statistica Sinica* **10**, 517–544.

Glickman, M. E., and Stern, H. S. (1998). A state-space model for National Football League scores. *Journal of the American Statistical Association* **93** 25–35.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106**, 746–762.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B* **69**, 243–268.

Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**, 359–378.

Goldstein, H. (1995). *Multilevel Statistical Models*, second edition. London: Edward Arnold.

Goldstein, H., and Silver, R. (1989). Multilevel and multivariate models in survey analysis. In *Analysis of Complex Surveys*, ed. C. J. Skinner, D. Holt, and T. M. F. Smith, 221–235. New York: Wiley.

Goldstein, M. (1976). Bayesian analysis of regression problems. *Biometrika* **63**, 51–58.

Golub, G. H., and van Loan, C. F. (1983). *Matrix Computations*. Baltimore: Johns Hopkins University Press.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. New York: Hafner.

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge, Mass.: MIT Press.

Goodman, L. A. (1952). Serial number analysis. *Journal of the American Statistical Association* **47**, 622–634.

Goodman, L. A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data (with discussion). *Journal of the American Statistical Association* **86**, 1085–1111.

Goodman, S. N. (1999a). Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine* **130**, 995–1013.

Goodman, S. N. (1999b). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130**, 1019–1021.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Greenland, S., Robins, J. M., and Pearl, J. (1999). Confounding and collapsability in causal inference. *Statistical Science* **14**, 29–46.

Greenland, S. (2001). Putting background information about relative risks into conjugate prior distributions. *Biometrics* **57**, 663–670.

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society A* **168**, 267–306.

Greenland, S., and Poole, C. (2013). Living with P-values: Resurrecting a Bayesian perspective on frequentist statistics (with discussion). *Epidemiology* **24**, 62–68.

Griewank, A., and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, second edition. Philadelphia: Society for Industrial and Applied Mathematics.

Griffin, J. E. (2011). The Ornstein-Uhlenbeck Dirichlet process and other time-varying processes for Bayesian nonparametric inference. *Journal of Statistical Planning and Inference* **141**, 3648–3664.

Griffin, J. E., and Steel, M. F. J. (2006). Order-based dependent Dirichlet process. *Journal of the American Statistical Association* **101**, 179–194.

Griffin, J. E., and Steel, M. F. J. (2011). Stick-breaking autoregressive processes. *Journal of Econometrics* **162**, 383–396.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A., eds. (2002). *Survey Nonresponse*. New York: Wiley.

Gull, S. F. (1989a). Developments in maximum entropy data analysis. In *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, 53–71. Dordrecht, Netherlands: Kluwer Academic Publishers.

Gull, S. F. (1989b). Bayesian data analysis: Straight-line fitting. In *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, 511–518. Dordrecht, Netherlands: Kluwer Academic Publishers.

Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society B* **29**, 83–100.

Hammersley, J. M., and Handscomb, D. C. (1964). *Monte Carlo Methods*. New York: Wiley.

Hannah, L., and Dunson, D. B. (2011). Bayesian nonparametric multivariate convex regression. http://arxiv.org/abs/1109.0322

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association* **99**, 609–619.

Hansen, M., and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association* **96**, 746–774.

Hanson, T., and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97**, 1020–1033.

Hartigan, J. (1964). Invariant prior distributions. *Annals of Mathematical Statistics* **35**, 836–845.

Hartley, H. O., and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

Harville, D. (1980). Predictions for NFL games with linear-model methodology. *Journal of the American Statistical Association* **75**, 516–524.

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models*. New York: Chapman & Hall.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hazelton, M. L., and Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. *Computational Statistics and Data Analysis* **55**, 2871–2879.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161.

Heinze, G., and Schemper, M. (2003). A solution to the problem of separation in logistic regression. *Statistics in Medicine* **12**, 2409–2419.

Heitjan, D. F. (1989). Inference from grouped continuous data: A review (with discussion). *Statistical Science* **4**, 164–183.

Heitjan, D. F., and Landis, J. R. (1994). Assessing secular trends in blood pressure: A multiple-imputation approach. *Journal of the American Statistical Association* **89**, 750–759.

Heitjan, D. F., Moskowitz, A. J., and Whang, W. (1999). Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics* **8**, 191–201.

Heitjan, D. F., and Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association* **85**, 304–314.

Heitjan, D. F., and Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics* **19**, 2244–2253.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association* **97**, 965–972.

Hensman, J., Fusi, N., Lawrence, N. (2013). Gaussian processes for big data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI 2013)*, ed. A. Nicholson and P. Smyth, 282–290.

Heskes, T., Opper, M., Wiegerinck, W., Winther, O., and Zoeter, O. (2005). Approximate inference techniques with expectation constraints. *Journal of Statistical Mechanics: Theory and Experiment*, P11015.

Hibbs, D. (2008). Implications of the 'bread and peace' model for the 2008 U.S. presidential election. *Public Choice* **137**, 1–10.

Higdon, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association* **93**, 585–595.

Higgins, J. P. T., and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**, 2733–2749.

Higgins, K. M., Davidian, M., Chew, G., and Burge, H. (1998). The effect of serial dilution error on calibration inference in immunoassay. *Biometrics* **54**, 19–32.

Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association* **60**, 806–825.

Hills, S. E., and Smith, A. F. M. (1992). Parameterization issues in Bayesian inference (with discussion). In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 227–246. Oxford University Press.

Hinde, J. (1982). Compound Poisson regression models. In *GLIM-82: Proceedings of the International Conference on Generalized Linear Models*, ed. R. Gilchrist (Lecture Notes in Statistics 14), 109–121. New York: Springer.

Hinkley, D. V., and Runger, G. (1984). The analysis of transformed data (with discussion). *Journal of the American Statistical Association* **79**, 302–320.

Hirano, K., Imbens, G., Rubin, D. B., and Zhao, X. H. (2000). Estimating the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics (with discussion). *Journal of the Royal Statistical Society B* **60**, 497–536.

Hodges, J. S., and Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly parameterized models. *Biometrika* **88**, 367–379.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.

Hoeting, J., Madigan, D., Raftery, A. E., and Volinsky, C. (1999). Bayesian model averaging (with discussion). *Statistical Science* **14**, 382–417.

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics* **1**, 265–283.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. New York: Springer.

Hoff, P. D., and Niu, X. (2012). A covariance regression model. *Statistica Sinica* **22**, 729–753.

Hoffman, M., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *The Journal of Machine Learning Research* **14**, 1303–1347.

Hoffman, M., and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*.

Hogan, H. (1992). The 1990 post-enumeration survey: An overview. *American Statistician* **46**, 261–269.

Hui, S. L., and Berger, J. O. (1983). Empirical Bayes estimation of rates in longitudinal studies. *Journal of the American Statistical Association* **78**, 753–760.

Imai, K., and van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*. **124**, 311–334.

Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706–710.

Imbens, G., and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–475.

Imbens, G., and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.

Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics* **17**, 295-311.

Ishwaran, H., and Zarepour, M. (2002). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica* **12**, 941–963.

Jaakkola, T. S., and Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing* **10**, 25–37.

Jackman, S. (2001). Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference and model checking. *Political Analysis* **9**, 227–241.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. New York: Wiley.

James, L. F., Lijoi, A., and Prunster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics* **36**, 76–97.

James, W., and Stein, C. (1960). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium* **1**, ed. J. Neyman, 361–380. Berkeley: University of California Press.

James, W. H. (1987). The human sex ratio. Part 1: A review of the literature. *Human Biology* **59**, 721–752.

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science* **20**, 50–67.

Jaynes, E. T. (1976). Confidence intervals vs. Bayesian intervals (with discussion). In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, ed. W. L. Harper and C. A. Hooker. Dordrecht, Netherlands: Reidel. Reprinted in Jaynes (1983).

Jaynes, E. T. (1980). Marginalization and prior probabilities. In *Bayesian Analysis in Econometrics and Statistics*, ed. A. Zellner, 43–87. Amsterdam: North-Holland. Reprinted in Jaynes (1983).

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE* **70**, 939–952.

Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*, ed. R. Rosenkrantz. Dordrecht, Netherlands: Reidel.

Jaynes, E. T. (1987). Bayesian spectrum and chirp analysis. In *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, ed. C. R. Smith and G. J. Erickson, 1–37. Dordrecht, Netherlands: Reidel.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffreys, H. (1961). *Theory of Probability*, third edition. Oxford University Press.

Joensuu, H., Reichardt, P., Eriksson, M., Hall, K. S., and Vehtari, A. (2014). Gastrointestinal stromal tumor: A method for optimizing the timing of CT scans in the follow-up of cancer patients. *Radiology* **271**, 96–106.

Joensuu, H., Vehtari, A., Riihimaki, J., Nishida, T., Steigen, S. E., Brabec, P., Plank, L., Nilsson, B., Cirilli, C., Braconi, C., Bordoni, A., Magnusson, M. K., Linke, Z., Sufliarsky, J., Massimo, F., Jonasson, J. G., Paolo Dei Tos, A., and Rutkowski, P. (2012). Risk of gastrointestinal stromal tumour recurrence after surgery: An analysis of pooled population-based cohorts. *Lancet Oncology* **13**, 265–274.

Johnson, N. L., and Kotz, S. (1972). *Distributions in Statistics*, 4 vols. New York: Wiley.

Johnson, V. E. (1996). On Bayesian analysis of multirater ordinal data: an application to automated essay grading. *Journal of the American Statistical Association* **91**, 42–51.

Johnson, V. E. (1997). An alternative to traditional GPA for evaluating student performance (with discussion). *Statistical Science* **12**, 251–278.

Johnson, V. E. (2004). A Bayesian χ^2 test for goodness-of-fit. *Annals of Statistics* **32**, 2361–2384.

Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L. (1999). Introduction to variational methods for graphical models. *Machine Learning* **37**, 183–233.

Jylanki, P., Nummenmaa, A., and Vehtari, A. (2013). Expectation propagation for neural networks with sparsity-promoting priors. *Journal of Machine Learning Research* **15**, 1849–1901.

Jylanki, P., Vanhatalo, J., and Vehtari, A. (2011). Robust Gaussian process regression with a Student-*t* likelihood. *Journal of Machine Learning Research* **12**, 3227–3257.

Kadane, J. B., and Seidenfeld, T. (1990). Randomization in a Bayesian perspective. *Journal of Statistical Planning and Inference* **25**, 329–345.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kahneman, D., and Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology* **3**, 430–454. Reprinted in *Judgment Under Uncertainty: Heuristics and Biases*, ed. Kahneman, D., Slovic, P., and Tversky, A., 32–47. Cambridge University Press (1982).

Karim, M. R., and Zeger, S. L. (1992). Generalized linear models with random effects; salamander mating revisited. *Biometrics* **48**, 631–644.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *American Statistician* **52**, 93–100.

Kass, R. E., and Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association* **90**, 773–795.

Kass, R. E., Tierney, L., and Kadane, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika* **76**, 663–674.

Kass, R. E., and Vaidyanathan, S. K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society B* **54**, 129–144.

Kass, R. E., and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association* **91**, 1343–1370.

Keller, J. B. (1986). The probability of heads. *American Mathematical Monthly* **93**, 191–197.

Kerman, J. (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics* **5**, 1450–1470.

Kerman, J., and Gelman, A. (2006). Bayesian data analysis using R. *R News* **6** (1), 21–24.

Kerman, J., and Gelman, A. (2007). Manipulating and summarizing posterior simulations using random variable objects. *Statistics and Computing* **17**, 235–244.

Kish, L. (1965). *Survey Sampling.* New York: Wiley.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* **5**, 1–25.

Kleinman, K. P., and Ibrahim, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics* **54**, 921–938.

Knuiman, M. W., and Speed, T. P. (1988). Incorporating prior information into the analysis of contingency tables. *Biometrics* **44**, 1061–1071.

Kong, A., Liu, J. S., and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association* **89**, 278–288.

Kong, A., McCullagh, P., Meng, X. L., Nicolae, D., and Tan, Z. (2003). A theory of statistical models for Monte Carlo integration (with discussion). *Journal of the Royal Statistical Society B* **65**, 585–618.

Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association* **94**, 1372–1381.

Kreft, I., and De Leeuw, J. (1998). *Introducing Multilevel Modeling.* London: Sage.

Kruschke, J. (2011). *Doing Bayesian Data Analysis.* New York: Academic Press.

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**, 76–86.

Kundu, S., and Dunson, D. B. (2014). Latent factor models for density estimation. *Biometrika* **101**, 641–654.

Kunsch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika* **74**, 517–524.

Laird, N. M., and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Landwehr, J. M., Pregibon, D., and Shoemaker, A. C. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association* **79**, 61–83.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881–896.

Lange, K., and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**, 175–198.

Laplace, P. S. (1785). Memoire sur les formules qui sont fonctions de tres grands nombres. In *Memoires de l'Academie Royale des Sciences.*

Laplace, P. S. (1810). Memoire sur les formules qui sont fonctions de tres grands nombres et sur leurs applications aux probabilites. In *Memoires de l'Academie des Sciences de Paris.*

Lau, J., Ioannidis, J. P. A., and Schmid, C. H. (1997). Quantitative synthesis in systematic reviews. *Annals of Internal Medicine* **127**, 820–826.

Lauritzen, S. L., and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society B* **50**, 157–224.

Lavine, M. (1991). Problems in extrapolation illustrated with space shuttle O-ring data (with discussion). *Journal of the American Statistical Association* **86**, 919–923.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *Annals of Statistics* **20**, 1222–1235.

Lax, J., and Phillips, J. (2009a). Gay rights in the states: Public opinion and policy responsiveness. *American Political Science Review* **103**, 367–386.

Lax, J., and Phillips, J. (2009b). How should we estimate public opinion in the states? *American Journal of Political Science* **53**, 107–121.

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1** (11), 277–330.

Le Cam, L., and Yang, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. New York: Springer.

Leamer, E. E. (1978a). Regression selection strategies and revealed priors. *Journal of the American Statistical Association* **73**, 580–587.

Leamer, E. E. (1978b). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Lee, P. M. (1989). *Bayesian Statistics: An Introduction*. Oxford University Press.

Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: Wiley.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, second edition. New York: Wiley.

Leimkuhler, B., and Reich, S. (2004). *Simulating Hamiltonian Dynamics*. Cambridge University Press.

Lenk, P. J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika* **78**, 531–543.

Lenk, P. J. (2003). Bayesian semiparametric density estimation and model verification using a logistic-Gaussian process. *Journal of Computational and Graphical Statistics* **12**, 548–565.

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika* **59**, 581–589.

Leonard, T. (1978). Density estimation, stochastic processes, and prior information. *Journal of the Royal Statistical Society B* **40**, 112–146.

Leonard, T., and Hsu, J. S. (1992). Bayesian inference for a covariance matrix. *Annals of Statistics* **20**, 1669–1696.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100**, 1989–2001.

Leyland, A. H., and Goldstein, H., eds. (2001). *Multilevel Modelling of Health Statistics*. Chichester: Wiley.

Liang, K. Y., and McCullagh, P. (1993). Case studies in binary dispersion. *Biometrics* **49**, 623–630.

Lin, C. Y., Gelman, A., Price, P. N., and Krantz, D. H. (1999). Analysis of local decisions using hierarchical modeling, applied to home radon measurement and remediation (with discussion). *Statistical Science* **14**, 305–337.

Lindgren, F., Rue, H., and Lindstrom, J. (2013). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society B* **73**, 423–498.

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society B* **20**, 102–107.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, two volumes. Cambridge University Press.

Lindley, D. V. (1971a). *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics.

Lindley, D. V. (1971b). The estimation of many parameters. In *Foundations of Statistical Science,* ed. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart and Winston.

Lindley, D. V., and Novick, M. R. (1981). The role of exchangeability in inference. *Annals of Statistics* **9**, 45–58.

Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–41.

Little, R. J. A. (1991). Inference with survey weights. *Journal of Official Statistics* **7**, 405–424.

Little, R. J. A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association* **88**, 1001–1012.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, second edition. New York: Wiley.

Liu, C. (1995). Missing data imputation using the multivariate t distribution. *Journal of Multivariate Analysis* **48**, 198–206.

Liu, C. (2004). Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X. L. Meng, 227–238. New York: Wiley.

Liu, C. (2003). Alternating subspace-spanning resampling to accelerate Markov chain Monte Carlo simulation. *Journal of the American Statistical Association* **98**, 110–117.

Liu, C., and Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–648.

Liu, C., and Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* **5**, 19–39.

Liu, C., Rubin, D. B., and Wu, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85**, 755–770.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Liu, J., and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**, 1264–1274.

Liu, Y., Gelman, A., Zheng, T., and Lee, D. (2013). Simulation-efficient shortest probability intervals. Technical report, Department of Statistics, Columbia University.

Lohr, S. (2009). *Sampling: Design and Analysis*, second edition. Pacific Grove, Calif.: Duxbury.

Longford, N. (1993). *Random Coefficient Models*. Oxford: Clarendon Press.

Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association* **78**, 393–398.

Louis, T. A., and Shen, W. (1999). Innovations in Bayes and empirical Bayes methods: estimating parameters, populations and ranks. *Statistics in Medicine* **18**, 2493–2505.

Luce, R. D., and Raiffa, H. (1957). *Games and Decisions*. New York: Wiley.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: evolution, critique and future directions (with discussion). *Statistics in Medicine* **28**, 3049–3082.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, 50–55.

MacEachern, S. N. (2000). Dependent nonparametric processes. Technical report, Department of Statistics, Ohio State University.

Madigan, D., and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association* **89**, 1535–1546.

Madow, W. G., Nisselson, H., Olkin, I., and Rubin, D. B. (1983). *Incomplete Data in Sample Surveys*, 3 vols. New York: Academic Press.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–675.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P., and Pellom, A. C. (1989). Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *Journal of the American Statistical Association* **84**, 637–650.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. New York: Academic Press.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* **22**, 1167–1180.

Marquardt, D. W., and Snee, R. D. (1975). Ridge regression in practice. *American Statistician* **29**, 3–19.

Marshall, E. C., and Spiegelhalter, D. J. (2007). Identifying outliers in Bayesian hierarchical models: A simulation-based approach. *Bayesian Analysis* **2**, 409–444.

Martin, A. D., and Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* **10**, 134–153.

Martz, H. F., and Zimmer, W. J. (1992). The risk of catastrophic failure of the solid rocket boosters on the space shuttle. *American Statistician* **46**, 42–47.

McClellan, M., McNeil, B. J., and Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction reduce mortality? *Journal of the American Medical Association* **272**, 859–866.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, second edition. New York: Chapman & Hall.

McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association* **84**, 473–478.

Meng, C. Y. K., and Dempster, A. P. (1987). A Bayesian approach to the multiplicity problem for significance testing with binomial data. *Biometrics* **43**, 301–311.

Meng, X. L. (1994a). On the rate of convergence of the ECM algorithm. *Annals of Statistics* **22**, 326–339.

Meng, X. L. (1994b). Multiple-imputation inferences with uncongenial sources of input (with discussion). *Statistical Science* **9**, 538–573.

Meng, X. L., and Pedlow, S. (1992). EM: A bibliographic review with missing articles. In *Proceedings of the American Statistical Association, Section on Statistical Computing*, 24–27.

Meng, X. L., Raghunathan, T. E., and Rubin, D. B. (1991). Significance levels from repeated p values with multiply-imputed data. *Statistica Sinica* **1**, 65–92.

Meng, X. L., and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association* **86**, 899–909.

Meng, X. L., and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply imputed data sets. *Biometrika* **79**, 103–111.

Meng, X. L., and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278.

Meng, X. L., and Schilling, S. (1996). Fitting full-information item factor models and empirical investigation of bridge sampling. *Journal of the American Statistical Association* **91**, 1254–1267.

Meng, X. L., and van Dyk, D. A. (1997). The EM algorithm—an old folk-song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society B* **59**, 511–567.

Meng, X. L., and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica* **6**, 831–860.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Metropolis, N., and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association* **44**, 335–341.

Meulders, M., Gelman, A., Van Mechelen, I., and De Boeck, P. (1998). Generalizing the probability matrix decomposition model: an example of Bayesian model checking and model expansion. In *Assumptions, Robustness, and Estimation Methods in Multivariate Modeling*, ed. J. Hox and E. D. de Leeuw, 1–19. Amsterdam: T-T Publikaties.

Minka, T. (2001). Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, ed. J. Breese and D. Koller, 362–369.

Mollie, A., and Richardson, S. (1991). Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine* **10**, 95–112.

Moody, J. E. (1992). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In *Advances in Neural Information Processing Systems 4*, ed. J. E. Moody, S. J. Hanson, and R. P. Lippmann, 847–854. San Francisco: Morgan Kaufmann Publishers.

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**, 284–289.

Moroff, S. V., and Pauker, S. G. (1983). What to do when the patient outlives the literature, or DEALE-ing with a full deck. *Medical Decision Making* **3**, 313–338.

Morris, C. (1983). Parametric empirical Bayes inference: theory and applications (with discussion). *Journal of the American Statistical Association* **78**, 47–65.

Mosteller, F., and Wallace, D. L. (1964). *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. New York: Springer. Reprinted 1984.

Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (2000). Fully model based approaches for spatially misaligned data. *Journal of the American Statistical Association* **95**, 877–887.

Muller, P., Quintana, F., and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society B* **66**, 735–749.

Muller, P., and Rosner, G. L. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92**, 1279–1292.

Mulligan, C. B., and Hunter, C. G. (2001). The empirical frequency of a pivotal vote. National Bureau of Economic Research Working Paper 8590.

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*.

Mykland, P., Tierney, L., and Yu, B. (1994). Regeneration in Markov chain samplers. *Journal of the American Statistical Association* **90**, 233–241.

Myllymaki, M., Sarkka, A., and Vehtari, A. (2013). Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates. *Spatial Statistics*.

Nandaram, B., and Sedransk, J. (1993). Bayesian predictive inference for a finite population proportion: Two-stage cluster sampling. *Journal of the Royal Statistical Society B* **55**, 399–408.

Neal, R. M. (1993). Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics* **111**, 194–203.

Neal, R. M. (1996a). *Bayesian Learning for Neural Networks*. New York: Springer.

Neal, R. M. (1996b). Sampling from multimodal distributions using tempered transitions. *Statistics and Computing* **6**, 353–366.

Neal, R. M. (1998). Regression and classification using Gaussian process priors (with discussion). In *Bayesian Statistics 6*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 475–501. Oxford University Press.

Neal, R. M. (2003). Slice sampling (with discussion). *Annals of Statistics* **31**, 705–767.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, ed. S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, 113–162. New York: Chapman & Hall.

Neelon, B., and Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60**, 398–406.

Nelder, J. A. (1977). A reformulation of linear models (with discussion). *Journal of the Royal Statistical Society A* **140**, 48–76.

Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statistics and Computing* **4**, 221–234.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* **135**, 370–384.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied Linear Statistical Models*, fourth edition. Burr Ridge, Ill.: Richard D. Irwin, Inc.

Newhouse, J. P., and McClellan, M. (1998). Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health* **19**, 17–34.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translated and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 463–480 (1990).

Normand, S. L., Glickman, M. E., and Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association* **92**, 803–814.

Normand, S. L., and Tritchler, D. (1992). Parameter updating in a Bayes network. *Journal of the American Statistical Association* **87**, 1109–1115.

Norvig, P. (2007). How to write a spelling corrector. `http://norvig.com/spell-correct.html`

Novick, M. R., Jackson, P. H., Thayer, D. T., and Cole, N. S. (1972). Estimating multiple regressions in m groups: A cross validation study. *British Journal of Mathematical and Statistical Psychology* **25**, 33–50.

Novick, M. R., Lewis, C., and Jackson, P. H. (1973). The estimation of proportions in m groups. *Psychometrika* **38**, 19–46.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society B* **40**, 1–42.

O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society B* **41**, 358–367.

O'Hagan, A. (1988). *Probability: Methods and Measurement.* New York: Chapman & Hall.

O'Hagan, A. (1991). Bayes-Hermite quadrature. *Journal of Statistical Planning and Inference* **29**, 245-260.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society B* **57**, 99–138.

O'Hagan, A. (2003). HSSS model criticism. In *Highly Structured Stochastic Systems*, ed. P. J. Green, N. L. Hjort, and S. Richardson, 423–444. Oxford University Press.

O'Hagan, A. (2004). Dicing with the unknown. *Significance* **1** (3), 132–133.

O'Hagan, A., and Forster, J. (2004). *Bayesian Inference*, second edition. London: Arnold.

Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2007). Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons. *Statistics in Medicine* **26**, 2088–2112.

Orchard, T., and Woodbury, M. A. (1972). A missing information principle: Theory and applications. In *Proceedings of the Sixth Berkeley Symposium*, ed. L. LeCam, J. Neyman, and E. L. Scott, 697–715. Berkeley: University of California Press.

Ormerod, J. T., and Wand, M. P. (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics* **21**, 2–17.

Ott, J. (1979). Maximum likelihood estimation by counting methods under polygenic and mixed models in human pedigrees. *American Journal of Human Genetics* **31**, 161–175.

Papaspiliopoulos, O., and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–186.

Pardoe, I. (2001). A Bayesian sampling approach to regression model checking. *Journal of Computational and Graphical Statistics* **10**, 617–627.

Pardoe, I., and Cook, R. D. (2002). A graphical method for assessing the fit of a logistic regression model. *American Statistician* **56**, 263–272.

Park, D., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis* **12**, 375–385.

Park, T., and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103**, 681–686.

Parmar, M. K. B., Griffiths, G. O., Spiegelhalter, D. J., Souhami, R. L., Altman, D. G., and van der Scheuren, E. (2001). Monitoring of large randomised clinical trials: A new approach with Bayesian methods. *Lancet* **358**, 375–381.

Parmigiani, G. (2002). *Modeling in Medical Decision Making: A Bayesian Approach.* New York: Wiley.

Parmigiani, G. (2004). Uncertainty and the value of diagnostic information. *Statistics in Medicine* **23**, 843–855.

Parmigiani, G., Berry, D., Iversen, E. S., Muller, P., Schildkraut, J., and Winer, E. (1999). Modeling risk of breast cancer and decisions about genetic testing (with discussion). In *Case Studies in Bayesian Statistics*, volume 4, ed. C. Gatsonis, R. E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, 133–203. New York: Springer.

Pati, D., and Dunson, D. B. (2011). Bayesian closed surface fitting through tensor products. Technical report, Department of Statistics, Duke University.

Pauler, D. K., Wakefield, J. C., and Kass, R. E. (1999). Bayes factors for variance component models. *Journal of the American Statistical Association* **94**, 1242–1253.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, Calif.: Morgan Kaufmann.

Pearl, J. (2010). *Causality*, second edition. Cambridge University Press.

Peltola, T., Marttinen, P., and Vehtari, A. (2012). Finite adaptation and multistep moves in the Metropolis-Hastings algorithm for variable selection in genome-wide association analysis. *PLoS One* **7**, e49445.

Peltonen, J., Venna, J., and Kaski, S. (2009). Visualizations for assessing convergence and mixing of Markov chain Monte Carlo simulations. *Computational Statistics and Data Analysis* **53**, 4453–4470.

Pericchi, L. R. (1981). A Bayesian approach to transformations to normality. *Biometrika* **68**, 35–43.

Pettitt, A. N., Friel, N., and Reeves, R. (2003). Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice. *Journal of the Royal Statistical Society B* **65**, 235–246.

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus.* New York: Springer.

Plackett, R. L. (1960). Models in the analysis of variance (with discussion). *Journal of the Royal Statistical Society B* **22**, 195–217.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Currently at `http://mcmc-jags.sourceforge.net`

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **9**, 523–539.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* New York: Chapman & Hall.

Polson, N. G., and Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics 9*, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, 501–539. Oxford University Press.

Polson, N. G., and Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* **7** (2), 1–16.

Pratt, J. W. (1965). Bayesian interpretation of standard inference statements (with discussion). *Journal of the Royal Statistical Society B* **27**, 169–203.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press.

Propp, J. G., and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9**, 223–252.

R Project (2002). The R project for statistical computing. `http://www.r-project.org`

Racine, A., Grieve, A. P., Fluhler, H., and Smith, A. F. M. (1986). Bayesian methods in practice: experiences in the pharmaceutical industry (with discussion). *Applied Statistics* **35**, 93–150.

Racine-Poon, A., Weihs, C., and Smith, A. F. M. (1991). Estimation of relative potency with sequential dilution errors in radioimmunoassay. *Biometrics* **47**, 1235–1246.

Raftery, A. E. (1988). Inference for the binomial N parameter: a hierarchical Bayes approach. *Biometrika* **75**, 223–228.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). In *Sociological Methodology 1995*, ed. P. V. Marsden.

Raftery, A. E. (1996a). Hypothesis testing and model selection via posterior simulation. In *Practical*

Markov Chain Monte Carlo, ed. W. Gilks, S. Richardson, and D. Spiegelhalter, 163–187. New York: Chapman & Hall.

Raftery, A. E. (1996b). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* **83**, 251–266.

Raghunathan, T. E. (1994). Monte Carlo methods for exploring sensitivity to distributional assumptions in a Bayesian analysis of a series of 2×2 tables. *Statistics in Medicine* **13**, 1525–1538.

Raghunathan, T. E., Lepkowski, J. E., Solenberger, P. W., and Van Hoewyk, J. H. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* **27**, 85–95.

Raghunathan, T. E., and Rubin, D. B. (1990). An application of Bayesian statistics using sampling/importance resampling for a deceptively simple problem in quality control. In *Data Quality Control: Theory and Pragmatics*, ed. G. Liepins and V. R. R. Uppuluri, 229–243. New York: Marcel Dekker.

Raiffa, H., and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Boston, Mass.: Harvard Business School.

Ramsay, J., and Silverman, B. W. (2005). *Functional Data Analysis*, second edition. New York: Springer.

Rasmussen, C. E., and Ghahramani, Z. (2003). Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems 15*, ed. S. Becker, S. Thrun, and K. Obermayer, 489–496. Cambridge, Mass.: MIT Press.

Rasmussen, C. E., and Nickish, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *Journal of Machine Learning Research* **11**, 3011–3015.

Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* Cambridge, Mass.: MIT Press.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models*, second edition. Thousand Oaks, Calif.: Sage.

Ray, S., and Mallick, B. (2006). Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society B* **68**, 305–332.

Reich, B. J., and Fuentes, M. (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics* **1**, 249–264.

Reilly, C., Gelman, A., and Katz, J. N. (2001). Post-stratification without population level information on the post-stratifying variable, with application to political polling. *Journal of the American Statistical Association* **96**, 1–11.

Reilly, C., and Zeringue, A. (2004). Improved predictions of lynx trappings using a biological model. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X. L. Meng, 297–308. New York: Wiley.

Ren, L., Du, L., Carin, L., and Dunson, D. B. (2011). Logistic stick-breaking processes. *Journal of Machine Learning Research* **12**, 203–239.

Richardson, S., and Gilks, W. R. (1993). A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology* **138**, 430–442.

Richardson, S., and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B* **59**, 731–792.

Riihimaki, J., Jylanki, P., and Vehtari, A. (2013). Nested expectation propagation for Gaussian process classification with a multinomial probit likelihood. *Journal of Machine Learning Research* **14**, 75–109.

Riihimaki, J., and Vehtari, A. (2010). Gaussian processes with monotonicity information. *Journal of Machine Learning Research: Workshop and Conference Proceedings* **9**, 645–652.

Riihimaki, J., and Vehtari, A. (2014). Laplace approximation for logistic Gaussian process density estimation and regression. *Bayesian Analysis* **9**, 425–448.

Ripley, B. D. (1981). *Spatial Statistics.* New York: Wiley.

Ripley, B. D. (1987). *Stochastic Simulation.* New York: Wiley.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

Robbins, H. (1955). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium* **1**, ed. J. Neyman, 157–164. Berkeley: University of California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Annals of Mathematical Statistics* **35**, 1–20.

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods*, second edition. New York: Springer.

Roberts, G. O., and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**, 351–367.

Roberts, G. O., and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society B* **59**, 291–317.

Robins, J. M. (1998). Confidence intervals for causal parameters. *Statistics in Medicine* **7**, 773–785.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* **6**, 15–51.

Rodriguez, A., and Dunson, D. B. (2011). Nonparametric Bayes models through probit stick-breaking processes. *Bayesian Analysis* **6**, 145–177.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* **103**, 1131–1144.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2009). Bayesian nonparametric functional data analysis through density estimation. *Biometrika* **96**, 149–162.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2010). Latent stick-breaking processes. *Journal of the American Statistical Association* **105**, 647–659.

Rodriguez, A., and ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis* **3**, 339–365.

Rombola, F. (1984). *The Book on Bookmaking*. Pasadena, Calif.: Pacific Book and Printing.

Romeel, D. (2011). Leapfrog integration.
`http://ursa.as.arizona.edu/~rad/phys305/ODE_III/node11.html`

Rosenbaum, P. R. (2010). *Observational Studies*, second edition. New York: Springer.

Rosenbaum, P. R., and Rubin, D. B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Rosenbaum, P. R., and Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society B* **45**, 212–218.

Rosenbaum, P. R., and Rubin, D. B. (1984a). Sensitivity of Bayes inference with data-dependent stopping rules. *American Statistician* **38**, 106–109.

Rosenbaum, P. R., and Rubin, D. B. (1984b). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* **79**, 516–524.

Rosenbaum, P. R., and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* **39**, 33–38.

Rosenkranz, S. L., and Raftery, A. E. (1994). Covariate selection in hierarchical models of hospital admission counts: a Bayes factor approach. Technical Report #268, Department of Statistics, University of Washington.

Rosenstone, S. (1984). *Forecasting Presidential Elections*. New Haven, Conn.: Yale University Press.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558–566.

Ross, S. M. (1983). *Stochastic Processes*. New York: Wiley.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1999). Adjusting for nonignorable dropout using semiparametric models. *Journal of the American Statistical Association* **94**, 1321–1339.

Rousseau, J., and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society B* **73**, 689–710.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57**, 377–387.

Rubin, D. B. (1974a). Characterizing the estimation of parameters in incomplete data problems. *Journal of the American Statistical Association* **69**, 467–474.

Rubin, D. B. (1974b). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* **2**, 1–26.

Rubin, D. B. (1978a). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **6**, 34–58.

Rubin, D. B. (1978b). Multiple imputations in sample surveys: a phenomenological Bayesian approach to nonresponse (with discussion). *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 20–34.

Rubin, D. B. (1980a). Discussion of 'Randomization analysis of experimental data: The Fisher randomization test,' by D. Basu. *Journal of the American Statistical Association* **75**, 591–593.

Rubin, D. B. (1980b). Using empirical Bayes techniques in the law school validity studies (with discussion). *Journal of the American Statistical Association* **75**, 801–827.

Rubin, D. B. (1981a). Estimation in parallel randomized experiments. *Journal of Educational Statistics* **6**, 377–401.

Rubin, D. B. (1981b). The Bayesian bootstrap. *Annals of Statistics* **9**, 130–134.

Rubin, D. B. (1983a). A case study of the robustness of Bayesian methods of inference: estimating the total in a finite population using transformations to normality. In *Scientific Inference, Data Analysis, and Robustness*, ed. G. E. P. Box, T. Leonard, and C. F. Wu, 213–244. New York: Academic Press.

Rubin, D. B. (1983b). Iteratively reweighted least squares. In *Encyclopedia of Statistical Sciences*, Vol. 4, ed. S. Kotz, N. L. Johnson, and C. B. Read, 272–275. New York: Wiley.

Rubin, D. B. (1983c). Progress report on project for multiple imputation of 1980 codes. Manuscript delivered to the U.S. Bureau of the Census, the U.S. National Science Foundation, and the Social Science Research Foundation.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12**, 1151–1172.

Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference. In *Bayesian Statistics 2*, ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, 463–472. Amsterdam: Elsevier Science Publishers.

Rubin, D. B. (1987a). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rubin, D. B. (1987b). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. Discussion of Tanner and Wong (1987). *Journal of the American Statistical Association* **82**, 543–546.

Rubin, D. B. (1989). A new perspective on meta-analysis. In *The Future of Meta-Analysis*, ed. K. W. Wachter and M. L. Straf. New York: Russell Sage Foundation.

Rubin, D. B. (1990). Discussion of 'On the application of probability theory to agricultural experiments. Essay on principles. Section 9,' by J. Neyman. *Statistical Science* **5**, 472–480.

Rubin, D. B. (1996). Multiple imputation after 18+ years (with discussion) *Journal of the American Statistical Association* **91**, 473–520.

Rubin, D. B. (1998). More powerful randomization-based p-values in double-blind trials with noncompliance (with discussion). *Statistics in Medicine* **17**, 371–385.

Rubin, D. B. (2000). Discussion of Dawid (2000). *Journal of the American Statistical Association* **95**, 435–438.

Rubin, D. B., and Schenker, N. (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological Methodology*, 131–144.

Rubin, D. B., and Stern, H. S. (1994). Testing in latent class models using a posterior predictive check distribution. In *Latent Variables Analysis: Applications for Developmental Research*, ed. A. Von Eye and C. C. Clogg, 420–438. Thousand Oaks, Calif.: Sage.

Rubin, D. B., Stern, H. S., and Vehovar, V. (1995). Handling 'Don't Know' survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association* **90**, 822–828.

Rubin, D. B., and Thomas, N. (1992). Affinely invariant matching methods with ellipsoidal distributions. *Annals of Statistics* **20**, 1079–93.

Rubin, D. B., and Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* **95**, 573–585.

Rubin, D. B., and Wu, Y. (1997). Modeling schizophrenic behavior using general mixture components. *Biometrics* **53**, 243–261.

Rue, H. (2013). The R-INLA project. `http://r-inla.org/`

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B* **71**, 319–382.

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918–924.

Sarkka, S. (2013). *Bayesian Filtering and Smoothing.* Cambridge University Press.

Sarkka, S., Solin, A., and Hartikainen, J. (2013). Spatio-temporal learning via infinite-dimensional Bayesian filtering and smoothing. *IEEE Signal Processing Magazine* **30**, 51–61.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110–114.

Savage, I. R. (1957). Nonparametric statistics. *Journal of the American Statistical Association* **52**, 331–344.

Savage, L. J. (1954). *The Foundations of Statistics.* New York: Dover.

Savitsky, T., Vannucci, M., and Sha, M. (2011). Variable selection for nonparametric Gaussian process priors: Models and computational strategies. *Statistical Science* **26**, 130–149.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* New York: Chapman & Hall.

Schmid, C. H., and Brown, E. N. (1999). A probability model for saltatory growth. In *Saltation and Stasis in Human Growth and Development: Evidence, Methods and Theory*, ed. M. Lampl, 121–131. London: Smith-Gordon.

Schmidt-Nielsen, K. (1984). *Scaling: Why is Animal Size So Important?* Cambridge University Press.

Schutt, R. (2009). Topics in model-based population inference. Ph.D. thesis, Department of Statistics, Columbia University.

Scott, A., and Smith, T. M. F. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association* **64**, 830–840.

Scott, A., and Smith, T. M. F. (1973). Survey design, symmetry and posterior distributions. *Journal of the Royal Statistical Society B* **55**, 57–60.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* New York: Wiley.

Seber, G. A. F. (1992). A review of estimating animal abundance II. *International Statistical Review* **60**, 129–166.

Sedlmeier, P., and Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General* **130**, 380-400.

Seeger, M. W. (2008). Bayesian inference and optimal design for the sparse linear model. *Journal of Machine Learning Research* **9**, 759–813.

Selvin, S. (1975). Letter. *American Statistician* **29**, 67.

Senn, S. (2013). Seven myths of randomisation in clinical trials. *Statistics in Medicine* **32**, 1439–1450.

Shafer, G. (1982). Lindley's paradox (with discussion). *Journal of the American Statistical Association* **77**, 325–351.

Sheiner, L. B., and Beal, S. L. (1982). Bayesian individualization of pharmacokinetics: Simple implementation and comparison with non-Bayesian methods. *Journal of Pharmaceutical Sciences* **71**, 1344–1348.

Sheiner, L. B., Rosenberg, B., and Melmon, K. L. (1972). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research* **5**, 441–459.

Shen, W., and Ghosal, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* **100**, 623–640.

Shen, W., and Louis, T. A. (1998). Triple-goal estimates in two-stage hierarchical models. *Journal of the Royal Statistical Society B* **60**, 455–471.

Shen, X., and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Annals of Statistics* **29**, 687–714.

Shibata, R. (1989). Statistical aspects of model selection. In *From Data to Model*, ed. J. C. Willems, 215–240. New York: Springer-Verlag.

Shirley, K., and Gelman, A. (2014). Hierarchical models for estimating state and demographic trends in U.S. death penalty public opinion. *Journal of the Royal Statistical Society A*.

Simoncelli, E. P. (1999). Bayesian denoising of visual images in the wavelet domain. In *Bayesian Inference in Wavelet Based Models*, ed. P. Muller and B. Vidakovic (Lecture Notes in Statistics 141), 291–308. New York: Springer.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The effects of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics* **15**, 217–230.

Sinharay, S., and Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference* **111**, 209–221.

Skare, O., Bolviken, E., and Holden, L. (2003). Improved sampling-importance resampling and reduced bias importance sampling. *Scandivanian Journal of Statistics* **30**, 719–737.

Skene, A. M., and Wakefield, J. C. (1990). Hierarchical models for multicentre binary response studies. *Statistics in Medicine* **9**, 910–929.

Skilling, J. (1989). Classic maximum entropy. In *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, 1–52. Dordrecht, Netherlands: Kluwer Academic Publishers.

Skinner, C. J., Holt, D., and Smith, T. M. F., eds. (1989). *The Analysis of Complex Surveys.* New York: Wiley.

Skrondal, A., and Rabe-Hesketh, S. (2014). Protective estimation of mixed-effects logistic regression when data are not missing at random. *Biometrika*

Smith, A. F. M. (1983). Bayesian approaches to outliers and robustness. In *Specifying Statistical Models from Parametric to Nonparametric, Using Bayesian or Non-Bayesian Approaches*, ed. J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith (Lecture Notes in Statistics 16), 13–35. New York: Springer.

Smith, A. F. M., and Gelfand, A. E. (1992). Bayesian statistics without tears. *American Statistician* **46**, 84–88.

Smith, A. F. M., and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B* **55**, 3–102.

Smith, A. F. M., Skene, A. M., Shaw, J. E. H., Naylor, J. C., and Dransfield, M. (1985). The implementation of the Bayesian paradigm. *Communications in Statistics* **14**, 1079–1102.

Smith, M., and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75**, 317–343.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* **14**, 2685–2699.

Smith, T. M. F. (1983). On the validity of inferences from non-random samples. *Journal of the Royal Statistical Society A* **146**, 394–403.

Snedecor, G. W., and Cochran, W. G. (1989). *Statistical Methods*, eighth edition. Ames: Iowa State University Press.

Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis.* London: Sage.

Snyder, J., with Herskowitz, M., and Perkins, S. (1975). *Jimmy the Greek, by Himself.* Chicago: Playboy Press.

Sommer, A., and Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.

Speed, T. P. (1990). Introductory remarks on Neyman (1923). *Statistical Science* **5**, 463–464.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B* **64**, 583–639.

Spiegelhalter, D. J., and Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society B* **44**, 377–387.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., and Lunn, D. (1994, 2003). BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England. `http://www.mrc-bsu.cam.ac.uk/bugs/`

Spitzer, E. (1999). The New York City Police Department's 'stop and frisk' practices. Office of the New York State Attorney General. `http://www.oag.state.ny.us/press/reports/stop_frisk/stop_frisk.html`

Stan Development Team (2012). Stan: A C++ library for probability and sampling. `http://mc-stan.org/`

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium* **1**, ed. J. Neyman, 197–206. Berkeley: University of California Press.

Stephens, M. (2000a). Bayesian analysis of mixture models with an unknown number of components: An alternative to reversible jump methods. *Annals of Statistics* **28**, 40–74.

Stephens, M. (2000b). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B* **62**, 795–809.

Stern, H. S. (1990). A continuum of paired comparison models. *Biometrika* **77**, 265–273.

Stern, H. S. (1991). On the probability of winning a football game. *American Statistician* **45**, 179–183.

Stern, H. S. (1997). How accurately can sports outcomes be predicted? *Chance* **10** (4), 19–23.

Stern, H. S. (1998). How accurate are the posted odds? *Chance* **11** (4), 17–21.

Sterne, J. A. C., and Smith, G. D. (2001). Sifting the evidence—what's wrong with significance tests? *British Medical Journal* **322**, 226–231.

Stigler, S. M. (1977). Do robust estimators work with real data? (with discussion). *Annals of Statistics* **5**, 1055–1098.

Stigler, S. M. (1983). Discussion of Morris (1983). *Journal of the American Statistical Association* **78**, 62–63.

Stigler, S. M. (1986). *The History of Statistics*. Cambridge, Mass.: Harvard University Press.

Stone, M. (1977). An asymptotic equivalence of choice of model cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society B* **36**, 44–47.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society B* **36**, 111–147.

Strenio, J. L. F., Weisberg, H. I., and Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics* **39**, 71–86.

Su, Y. S., Gelman, A., Hill, J., and Yajima, M. (2011). Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *Journal of Statistical Software* **45** (2).

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, third edition. New York: Springer.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550.

Taplin, R. H., and Raftery, A. E. (1994). Analysis of agricultural field trials in the presence of outliers and fertility jumps. *Biometrics* **50**, 764–781.

Tarone, R. E. (1982). The use of historical control information in testing for a trend in proportions. *Biometrics* **38**, 215–220.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**, 1566–1581.

Thall, P. F., Simon, R. M., and Estey, E. H. (1995). Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine* **14**, 357–379.

Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H., and Benjamin, R. S. (2003). Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Statistics in Medicine* **22**, 763–780.

Thisted, R. (1988). *Elements of Statistical Computing: Numerical Computation*. New York: Chapman & Hall.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 837–842. Oxford University Press.

Tiao, G. C., and Box, G. E. P. (1967). Bayesian analysis of a three-component hierarchical design model. *Biometrika* **54**, 109–125.

Tiao, G. C., and Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I: Posterior distribution of variance components. *Biometrika* **52**, 37–53.

Tiao, G. C., and Tan, W. Y. (1966). Bayesian analysis of random-effect models in the analysis of variance. II: Effect of autocorrelated errors. *Biometrika* **53**, 477–495.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* **58**, 267–288.

Tibshirani, R. J., and Tibshirani, R. (2009). A bias correction for the minimum error rate in cross-validation. *Annals of Applied Statistics* **3**, 822–829.

Tierney, L., and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86.

Tierney, L. (1998). A note on the Metropolis Hastings algorithm for general state spaces. *Annals of Applied Probability* **8**, 1–9.

Tipping, M. E., and Lawrence, N. D. (2005). Variational inference for Student-t models: Robust Bayesian interpolation and generalised component analysis. *Neurocomputing* **69**, 123–141.

Titterington, D. M. (1984). The maximum entropy method for data analysis (with discussion). *Nature* **312**, 381–382.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

Tokdar, S. T. (2007). Towards a faster implementation of density estimation with logistic Gaussian process priors. *Journal of Computational and Graphical Statistics* **16**, 633–655.

Tokdar, S. T. (2011). Adaptive convergence rates of a Dirichlet process mixture of multivariate normals. `http://arxiv.org/abs/1111.4148`

Tokdar, S. T., and Ghosh, J. K. (2007). Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference* **137**, 34–42.

Tokdar, S. T., Zhu, Y. M., and Ghosh, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Analysis* **5**, 319–344.

Tolvanen, V., Jylanki, P., and Vehtari, A. (2014). Expectation propagation for nonstationary heteroscedastic Gaussian process regression. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, DOI:10.1109/MLSP.2014.6958906.

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., and Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices. Technical report, Department of Psychology, University of Leuven.

Tsui, K. W., and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association* **84**, 602–607.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, Conn.: Graphics Press.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, Conn.: Graphics Press.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

Turner, D. A., and West, M. (1993). Bayesian analysis of mixtures applied to postsynaptic potential fluctuations. *Journal of Neuroscience Methods* **47**, 1–23.

Vaida, F., and Blanchard, S. (2002). Conditional Akaike information for mixed effects models. Technical report, Department of Biostatistics, Harvard University.

Vail, A., Hornbuckle, J., Spiegelhalter, D. J., and Thomas, J. G. (2001). Prospective application of Bayesian monitoring and analysis in an 'open' randomized clinical trial. *Statistics in Medicine* **20**, 3777–3787.

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. London: Chapman & Hall.

Van Buuren, S., Boshuizen, H. C., and Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* **18**, 681–694.

Van Buuren, S., and Oudshoom, C. G. M. (2000). MICE: Multivariate imputation by chained equations (S software for missing-data imputation). `http://web.inter.nl.net/users/S.van.Buuren/mi/`

van der Linde, A. (2005). DIC in variable selection. *Statistica Neerlandica* **59**, 45–56.

van Dyk, D. A., and Meng, X. L. (2001). The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics* **10**, 1–111.

van Dyk, D. A., Meng, X. L., and Rubin, D. B. (1995). Maximum likelihood estimation via the ECM algorithm: computing the asymptotic variance. *Statistica Sinica* **5**, 55–75.

Vanhatalo, J., Jylanki, P., and Vehtari, A. (2009). Gaussian process regression with Student-t likelihood. *Advances in Neural Information Processing Systems 22*, ed. Y. Bengio et al., 1910–1918.

Vanhatalo, J., Pietilainen, V., and Vehtari, A. (2010). Approximate inference for disease mapping with sparse Gaussian processes. *Statistics in Medicine* **29**, 1580–1607.

Vanhatalo, J., Riihimaki, J., Hartikainen, J., Jylanki, P., Tolvanen, V., and Vehtari, A. (2013a). Bayesian modeling with Gaussian processes using the GPstuff toolbox. `http://arxiv.org/abs/1206.5754`

Vanhatalo, J., Riihimaki, J., Hartikainen, J., Jylanki, P., Tolvanen, V., and Vehtari, A. (2013b). GPstuff: Bayesian modeling with Gaussian processes. *Journal of Machine Learning Research* **14**, 1005–1009.

Vanhatalo, J., and Vehtari, A. (2010). Speeding up the binary Gaussian process classification. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, ed. P. Grunwald and P. Spirtes, 623–632.

Vehtari, A., and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation* **14**, 2439–2468.

Vehtari, A., and Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys* **6**, 142–228.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*, fourth edition. New York: Springer.

Venna, J., Kaski, S., and Peltonen, J. (2003). Visualizations for assessing convergence and mixing of MCMC. In *Machine Learning: ECML 2003, Lecture Notes in Artificial Intelligence, Vol. 2837*, ed. N. Lavrae, D. Gamberger, H. Blockeel, and L. Todorovski. Berlin: Springer.

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Volfovsky, A., and Hoff, P. (2012). Hierarchical array priors for ANOVA decompositions. Technical report, Department of Statistics, University of Washington.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society B* **40**, 364–372.

Wainer, H. (1984). How to display data badly. *American Statistician* **38**, 137–147.

Wainer, H. (1997). *Visual Revelations*. New York: Springer.

Wakefield, J. C. (1996). The Bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association* **91**, 62–75.

Wakefield, J. C., Aarons, L., and Racine-Poon, A. (1999). The Bayesian approach to population pharmacokinetic/pharmacodynamic modeling (with discussion). In *Case Studies in Bayesian Statistics*, volume 4, ed. C. Gatsonis, R. E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, 205–265. New York: Springer.

Wakefield, J. C., Gelfand, A. E., and Smith, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing* **1**, 129–133.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607–617.

Waller, L. A., Louis, T. A., and Carlin, B. P. (1997). Bayes methods for combining disease and exposure data in assessing environmental justice. *Environmental and Ecological Statistics* **4**, 267–281.

Wang, L., and Dunson, D. B. (2011a). Fast Bayesian inference in Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **20**, 196–216.

Wang, L., and Dunson, D. B. (2011b). Bayesian isotonic density regression. *Biometrika* **98**, 537–551.

Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference (with discussion). In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 438–502. Oxford University Press.

Wasserman, L. (2000). Asymptotic inference for mixture models using data dependent priors. *Journal of the Royal Statistical Society B* **62**, 159–180.

Watanabe, S. (2009). *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research* **14**, 867–897.

Weisberg, S. (1985). *Applied Linear Regression*, second edition. New York: Wiley.

Weiss, R. E. (1994). Pediatric pain, predictive inference, and sensitivity analysis. *Evaluation Review* **18**, 651–678.

Weiss, R. E. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society B* **58**, 739–750.

Wermuth, N., and Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs, and graphical chain models. *Journal of the Royal Statistical Society B* **52**, 21–50.

West, M. (1992). Modelling with mixtures. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 503–524. Oxford University Press.

West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In *Bayesian Statistics 7*, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, 733–742. Oxford University Press.

West, M., and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96**, 382–397.

Wong, F., Carter, C., and Kohn, R. (2002). Efficient estimation of covariance selection models. Technical report, Australian Graduate School of Management.

Wong, W. H., and Li, B. (1992). Laplace expansion for posterior densities of nonlinear functions of parameters. *Biometrika* **79**, 393–398.

Yang, R., and Berger, J. O. (1994). Estimation of a covariance matrix using reference prior. *Annals of Statistics* **22**, 1195–1211.

Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **4**, 777–790.

Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases* **27**, 335–371.

Zaslavsky, A. M. (1993). Combining census, dual-system, and evaluation study data to estimate population shares. *Journal of the American Statistical Association* **88**, 1092–1105.

Zeger, S. L., and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86.

Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine* **300**, 1242–1245.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zellner, A. (1975). Bayesian analysis of regression error terms. *Journal of the American Statistical Association* **70**, 138–144.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *Journal of the American Statistical Association* **71**, 400–405.

Zhang, J. (2002). Causal inference with principal stratification: Some theory and application. Ph.D. thesis, Department of Statistics, Harvard University.

Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Annals of Statistics* **28**, 532–552.

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis* **13**, 157–170.

This page intentionally left blank

This page intentionally left blank

This page intentionally left blank

Statistics

Bayesian Data Analysis

Third Edition

Now in its third edition, this classic book is widely considered the leading text on Bayesian methods, lauded for its accessible, practical approach to analyzing data and solving research problems. **Bayesian Data Analysis, Third Edition** continues to take an applied approach to analysis using up-to-date Bayesian methods. The authors—all leaders in the statistics community—introduce basic concepts from a data-analytic perspective before presenting advanced methods. Throughout the text, numerous worked examples drawn from real applications and research emphasize the use of Bayesian inference in practice.

New to the Third Edition

- Four new chapters on nonparametric modeling
- Coverage of weakly informative priors and boundary-avoiding priors
- Updated discussion of cross-validation and predictive information criteria
- Improved convergence monitoring and effective sample size calculations for iterative simulation
- Presentations of Hamiltonian Monte Carlo, variational Bayes, and expectation propagation
- New and revised software code

Suitable for both students and researchers, the book introduces Bayesian inference starting from first principles, presents effective current approaches to Bayesian modeling and computation in statistics and related fields, and provides an assortment of Bayesian methods in applied statistics. Additional materials, including data sets used in the examples, are available on the book's web page.

K11900

ISBN: 978-1-4398-4095-5

90000

9 781439 840955

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK