Bayesian Data Analysis
Third edition
(Draft, 15 July 2013)

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

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.

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, 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, 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 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.