

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

Copyright ©1995–2013, by Andrew Gelman, John Carlin, Hal Stern,
Donald Rubin, David Dunson, and Aki Vehtari



Contents

Preface	xiii
Part I: Fundamentals of Bayesian Inference	1
1 Probability and inference	3
1.1 The three steps of Bayesian data analysis	3
1.2 General notation for statistical inference	4
1.3 Bayesian inference	6
1.4 Discrete probability examples: genetics and spell checking	8
1.5 Probability as a measure of uncertainty	11
1.6 Example of probability assignment: football point spreads	13
1.7 Example: estimating the accuracy of record linkage	16
1.8 Some useful results from probability theory	19
1.9 Computation and software	22
1.10 Bayesian inference in applied statistics	24
1.11 Bibliographic note	25
1.12 Exercises	27
2 Single-parameter models	29
2.1 Estimating a probability from binomial data	29
2.2 Posterior as compromise between data and prior information	32
2.3 Summarizing posterior inference	32
2.4 Informative prior distributions	34
2.5 Estimating a normal mean with known variance	39
2.6 Other standard single-parameter models	42
2.7 Example: informative prior distribution for cancer rates	47
2.8 Noninformative prior distributions	51
2.9 Weakly informative prior distributions	55
2.10 Bibliographic note	56
2.11 Exercises	57
3 Introduction to multiparameter models	63
3.1 Averaging over ‘nuisance parameters’	63
3.2 Normal data with a noninformative prior distribution	64
3.3 Normal data with a conjugate prior distribution	67
3.4 Multinomial model for categorical data	69
3.5 Multivariate normal model with known variance	70
3.6 Multivariate normal with unknown mean and variance	72
3.7 Example: analysis of a bioassay experiment	74
3.8 Summary of elementary modeling and computation	78
3.9 Bibliographic note	78
3.10 Exercises	79

4	Asymptotics and connections to non-Bayesian approaches	83
4.1	Normal approximations to the posterior distribution	83
4.2	Large-sample theory	87
4.3	Counterexamples to the theorems	89
4.4	Frequency evaluations of Bayesian inferences	91
4.5	Bayesian interpretations of other statistical methods	92
4.6	Bibliographic note	97
4.7	Exercises	98
5	Hierarchical models	101
5.1	Constructing a parameterized prior distribution	102
5.2	Exchangeability and setting up hierarchical models	104
5.3	Fully Bayesian analysis of conjugate hierarchical models	108
5.4	Estimating exchangeable parameters from a normal model	113
5.5	Example: parallel experiments in eight schools	118
5.6	Hierarchical modeling applied to a meta-analysis	123
5.7	Weakly informative priors for hierarchical variance parameters	128
5.8	Bibliographic note	132
5.9	Exercises	134
	Part II: Fundamentals of Bayesian Data Analysis	139
6	Model checking	141
6.1	The place of model checking in applied Bayesian statistics	141
6.2	Do the inferences from the model make sense?	142
6.3	Posterior predictive checking	143
6.4	Graphical posterior predictive checks	153
6.5	Model checking for the educational testing example	159
6.6	Bibliographic note	161
6.7	Exercises	163
7	Evaluating, comparing, and expanding models	165
7.1	Measures of predictive accuracy	166
7.2	Information criteria and cross-validation	169
7.3	Model comparison based on predictive performance	178
7.4	Model comparison using Bayes factors	182
7.5	Continuous model expansion	184
7.6	Implicit assumptions and model expansion: an example	187
7.7	Bibliographic note	192
7.8	Exercises	193
8	Modeling accounting for data collection	197
8.1	Bayesian inference requires a model for data collection	197
8.2	Data-collection models and ignorability	199
8.3	Sample surveys	205
8.4	Designed experiments	214
8.5	Sensitivity and the role of randomization	218
8.6	Observational studies	220
8.7	Censoring and truncation	224
8.8	Discussion	229
8.9	Bibliographic note	229
8.10	Exercises	230

9 Decision analysis	237
9.1 Bayesian decision theory in different contexts	238
9.2 Using regression predictions: incentives for telephone surveys	239
9.3 Multistage decision making: medical screening	245
9.4 Hierarchical decision analysis for radon measurement	246
9.5 Personal vs. institutional decision analysis	255
9.6 Bibliographic note	257
9.7 Exercises	257
Part III: Advanced Computation	259
10 Introduction to Bayesian computation	261
10.1 Numerical integration	261
10.2 Distributional approximations	262
10.3 Direct simulation and rejection sampling	263
10.4 Importance sampling	265
10.5 How many simulation draws are needed?	267
10.6 Computing environments	268
10.7 Debugging Bayesian computing	270
10.8 Bibliographic note	271
10.9 Exercises	272
11 Basics of Markov chain simulation	275
11.1 Gibbs sampler	276
11.2 Metropolis and Metropolis-Hastings algorithms	278
11.3 Using Gibbs and Metropolis as building blocks	280
11.4 Inference and assessing convergence	281
11.5 Effective number of simulation draws	286
11.6 Example: hierarchical normal model	288
11.7 Bibliographic note	291
11.8 Exercises	291
12 Computationally efficient Markov chain simulation	293
12.1 Efficient Gibbs samplers	293
12.2 Efficient Metropolis jumping rules	295
12.3 Further extensions to Gibbs and Metropolis	297
12.4 Hamiltonian Monte Carlo	300
12.5 Hamiltonian dynamics for a simple hierarchical model	305
12.6 Stan: developing a computing environment	307
12.7 Bibliographic note	308
12.8 Exercises	309
13 Modal and distributional approximations	311
13.1 Finding posterior modes	311
13.2 Boundary-avoiding priors for modal summaries	313
13.3 Normal and related mixture approximations	318
13.4 Finding marginal posterior modes using EM	320
13.5 Approximating conditional and marginal posterior densities	325
13.6 Example: hierarchical normal model (continued)	326
13.7 Variational inference	331
13.8 Expectation propagation	338
13.9 Other approximations	343

13.10 Unknown normalizing factors	345
13.11 Bibliographic note	348
13.12 Exercises	349
Part IV: Regression models	353
14 Introduction to regression models	355
14.1 Conditional modeling	355
14.2 Bayesian analysis of the classical regression model	356
14.3 Example: incumbency in congressional elections	360
14.4 Goals of regression analysis	366
14.5 Assembling the matrix of explanatory variables	367
14.6 Regularization and dimension reduction for multiple predictors	370
14.7 Unequal variances and correlations	371
14.8 Including numerical prior information	378
14.9 Bibliographic note	380
14.10 Exercises	380
15 Hierarchical linear models	385
15.1 Regression coefficients exchangeable in batches	386
15.2 Example: forecasting U.S. presidential elections	387
15.3 Interpreting a normal prior distribution as additional data	392
15.4 Varying intercepts and slopes	394
15.5 Computation: batching and transformation	396
15.6 Analysis of variance and the batching of coefficients	399
15.7 Hierarchical models for batches of variance components	402
15.8 Bibliographic note	404
15.9 Exercises	405
16 Generalized linear models	409
16.1 Standard generalized linear model likelihoods	410
16.2 Working with generalized linear models	411
16.3 Weakly informative priors for logistic regression	416
16.4 Example: hierarchical Poisson regression for police stops	424
16.5 Example: hierarchical logistic regression for political opinions	426
16.6 Models for multivariate and multinomial responses	427
16.7 Loglinear models for multivariate discrete data	432
16.8 Bibliographic note	435
16.9 Exercises	436
17 Models for robust inference	439
17.1 Aspects of robustness	439
17.2 Overdispersed versions of standard probability models	441
17.3 Posterior inference and computation	443
17.4 Robust inference and sensitivity analysis for the 8 schools	445
17.5 Robust regression using t -distributed errors	448
17.6 Bibliographic note	449
17.7 Exercises	450

CONTENTS	xi
18 Models for missing data	453
18.1 Notation	453
18.2 Multiple imputation	455
18.3 Missing data in the multivariate normal and t models	458
18.4 Example: multiple imputation for a series of polls	460
18.5 Missing values with counted data	466
18.6 Example: an opinion poll in Slovenia	467
18.7 Bibliographic note	470
18.8 Exercises	471
Part V: Nonlinear and nonparametric models	473
19 Parametric nonlinear models	475
19.1 Example: serial dilution assay	475
19.2 Example: population toxicokinetics	481
19.3 Bibliographic note	489
19.4 Exercises	489
20 Basis function models	491
20.1 Splines and weighted sums of basis functions	491
20.2 Basis selection and shrinkage of coefficients	494
20.3 Non-normal models and multivariate regression surfaces	498
20.4 Bibliographic note	502
20.5 Exercises	502
21 Gaussian process models	505
21.1 Gaussian process regression	505
21.2 Example: birthdays and birthdates	509
21.3 Latent Gaussian process models	513
21.4 Functional data analysis	516
21.5 Density estimation and regression	516
21.6 Bibliographic note	519
21.7 Exercises	520
22 Finite mixture models	523
22.1 Setting up and interpreting mixture models	523
22.2 Example: reaction times and schizophrenia	528
22.3 Label switching and posterior computation	537
22.4 Unspecified number of mixture components	540
22.5 Mixture models for classification and regression	543
22.6 Bibliographic note	546
22.7 Exercises	547
23 Dirichlet process models	549
23.1 Bayesian histograms	549
23.2 Dirichlet process prior distributions	550
23.3 Dirichlet process mixtures	553
23.4 Beyond density estimation	561
23.5 Hierarchical dependence	564
23.6 Density regression	572
23.7 Bibliographic note	575
23.8 Exercises	577

Appendixes	579
A Standard probability distributions	581
A.1 Continuous distributions	581
A.2 Discrete distributions	589
A.3 Bibliographic note	590
B Outline of proofs of limit theorems	591
B.1 Bibliographic note	594
C Computation in R and Stan	595
C.1 Getting started with R and Stan	595
C.2 Fitting a hierarchical model in Stan	595
C.3 Direct simulation, Gibbs, and Metropolis in R	600
C.4 Programming Hamiltonian Monte Carlo in R	607
C.5 Further comments on computation	611
C.6 Bibliographic note	612

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 McPeck, Xiao-Li Meng, Baback Moghadam, 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.

¹In 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.