CRC Press
Taylor & Francis Group

Chapman & Hall/CRC Handbooks of Modern Statistical Methods: Contributors

Distribution
If a library purchases our e-book package, all people on the network have free access so a large number of readers and downloads are likely.

Formatting
LaTeX: Use the Krantz2 style file. The text area is 5.5 x 8.5 inches.

Word: Give us an unformatted file; do not try to make the article look like the sample pages. Cancel the automatic hyphenation feature and do not use end-of-line hyphens. Do not right-justify. To estimate book pages, assume 500 words/page. CRC will be responsible for the final formatting and indexing.

Abstracts
All articles should include abstracts of 150–200 words plus an ORCID iD available at https://orcid.org. LaTeX authors should supply them as a separate text file; Word authors should include them on the chapter opening page below the title and contributor names.

Figures
Almost all handbooks are printed in grayscale. It is okay to include color in figures if they will look okay in grayscale. **Do not use color in the legend as the only difference between lines.**

Word: Figures should be provided as individual files labeled with the correct figure number. Tables should be placed at the end of the contribution. Include in-text notations such as "Insert Figure 1.1 here." for figures and tables. Include figure/table captions at the end of each chapter.

Equations
Word: Equations can be written with the Word equation editor or with MathType.

Free copies
Each co-contributor receives a free print or e-book copy.

Front matter
Contributors should supply the affiliation to be included in the front matter (p. 7). The first page of the article includes only your name (p. 8).

References
Type references in the style of the sample pages. (pp. 9–10).

Index
LaTeX: Index your article; Word: Index it with Word or send us a list of words for us to index.

Permission forms and contributor authorization agreement
You must sign a contributor's agreement at an early stage for your article to be included. All articles must be original although they can be based on more technical journal articles.

LaTeX help
Help with our style file or LaTeX issues is available at texhelp@taylorandfrancis.com.

Contents

Preface

Over the past 20 years or so, Markov Chain Monte Carlo (MCMC) methods have revolutionized statistical computing. They have impacted the practice of Bayesian statistics profoundly by allowing intricate models to be posited and *used* in an astonishing array of disciplines as diverse as fisheries science and economics. Of course, Bayesians are not the only ones to benefit from using MCMC, and there continues to be increasing use of MCMC in other statistical settings. The practical importance of MCMC has also sparked expansive and deep investigation into fundamental Markov chain theory. As the use of MCMC methods mature, we see deeper theoretical questions addressed, more complex applications undertaken and their use spreading to new fields of study. It seemed to us that it was a good time to try to collect an overview of MCMC research and its applications.

This book is intended to be a reference (not a text) for a broad audience and to be of use both to developers and users of MCMC methodology. There is enough introductory material in the book to help graduate students as well as researchers new to MCMC who wish to become acquainted with the basic theory, algorithms and applications. The book should also be of particular interest to those involved in the development or application of new and advanced MCMC methods. Given the diversity of disciplines that use MCMC, it seemed prudent to have many of the chapters devoted to detailed examples and case studies of realistic scientific problems. Those wanting to see current practice in MCMC will find a wealth of material to choose from here.

Roughly speaking, we can divide the book into two parts. The first part encompasses 12 chapters concerning MCMC foundations, methodology and algorithms. The second part consists of 12 chapters which consider the use of MCMC in practical applications. Within the first part, the authors take such a wide variety of approaches that it seems pointless to try to classify the chapters into subgroups. For example, some chapters attempt to appeal to a broad audience by taking a tutorial approach while other chapters, even if introductory, are either more specialized or present more advanced material. Yet others present original research. In the second part, the focus shifts to applications. Here again, we see a variety of topics, but there are two basic approaches taken by the authors of these chapters. The first is to provide an overview of an application area with the goal of identifying best MCMC practice in the area through extended examples. The second approach is to provide detailed case studies of a given problem while clearly identifying the statistical and MCMC-related issues encountered in the application.

When we were planning this book, we quickly realized that no single source can give a truly comprehensive overview of cutting-edge MCMC research and applications—there is just too much of it and its development is moving too fast. Instead, the editorial goal was to obtain contributions of high quality that may stand the test of time. To this end, all of the contributions (including those written by members of the editorial panel) were submitted to a rigorous peer review process and many underwent several revisions. Some contributions, even after revisions, were deemed unacceptable for publication here, and we certainly welcome constructive feedback on the chapters that did survive our editorial process. We thank all the authors for their efforts and patience in this process, and we ask for understanding from those whose contributions are not included in this book. We believe the breadth and depth of the contributions to this book, including some diverse opinions expressed, imply a continuously bright and dynamic future for MCMC research. We hope this book inspires further work—theoretical, methodological, and applied—in this exciting and rich area.

Finally, no project of this magnitude could be completed with satisfactory outcome without many individuals' help. We especially want to thank Robert Calver of Chapman & Hall/CRC for his encouragements, guidelines, and particularly his patience during the entire process of editing this book. We also offer our heartfelt thanks to the numerous referees for their insightful and rigorous review, often multiple times. Of course, the ultimate appreciation for all individuals involved in this project comes from your satisfaction with the book or at least a part of it. So we thank you for reading it.

Steve Brooks
Andrew Gelman
Galin L. Jones
Xiao-Li Meng

Editors

Steve Brooks is company director of ATASS, a statistical consultancy business based in the United Kingdom. He was formerly professor of Statistics at Cambridge University and received the Royal Statistical Society Guy medal in Bronze in 2005 and the Philip Leverhulme prize in 2004. Like his co-editors, he has served on numerous professional committees both in the United Kingdom and elsewhere, as well as sitting on numerous editorial boards. He is co-author of *Bayesian Analysis for Population Ecology* (Chapman & Hall/CRC, 2009) and co-founder of the National Centre for Statistical Ecology. His research interests include the development and application of computational statistical methodology across a broad range of application areas.

Andrew Gelman is a professor of statistics and political science and director of the Applied Statistics Center at Columbia University. He has received the Outstanding Statistical Application award from the American Statistical Association, the award for best article published in the *American Political Science Review*, and the Committee of Presidents of Statistical Societies award for outstanding contributions by a person under the age of 40. His books include *Bayesian Data Analysis* (with John Carlin, Hal Stern, and Don Rubin), *Teaching Statistics: A Bag of Tricks* (with Deb Nolan), *Data Analysis Using Regression and Multilevel/Hierarchical Models* (with Jennifer Hill), and, most recently, *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do* (with David Park, Boris Shor, Joe Bafumi, and Jeronimo Cortina).

Andrew has done research on a wide range of topics, including: why it is rational to vote; why campaign polls are so variable when elections are so predictable; why redistricting is good for democracy; reversals of death sentences; police stops in New York City; the statistical challenges of estimating small effects; the probability that your vote will be decisive; seats and votes in Congress; social network structure; arsenic in Bangladesh; radon in your basement; toxicology; medical imaging; and methods in surveys, experimental design, statistical inference, computation, and graphics.

Galin L. Jones is an associate professor in the School of Statistics at the University of Minnesota. He has served on many professional committees and is currently serving on the editorial board for the *Journal of Computational and Graphical Statistics*. His research interests include Markov chain Monte Carlo, Markov chains in decision theory, and applications of statistical methodology in agricultural, biological, and environmental settings.

Xiao-Li Meng is the Whipple V. N. Jones professor of statistics and chair of the Department of Statistics at Harvard University; previously he taught at the University of Chicago (1991–2001). He was the recipient of the 1997–1998 University of Chicago Faculty Award for Excellence in Graduate Teaching, the 2001 Committee of Presidents of Statistical Societies Award, the 2003 Distinguished Achievement Award and the 2008 Distinguished Service Award from the International Chinese Statistical Association, and the 2010 Medallion Lecturer from the Institute of Mathematical Statistics (IMS). He has served on numerous professional committees, including chairing the 2004 Joint Statistical Meetings and the Committee on Meetings of American Statistical Association (ASA) from 2004 until 2010. He is an elected fellow of the ASA and the IMS. He has also served on editorial boards for *The Annals of Statistics*, *Bayesian Analysis*, *Bernoulli*, *Biometrika*, *Journal of the American Statistical*

Contributors

Susan Spear Bassett
Department of Psychiatry and
Behavioral Sciences
Johns Hopkins Hospital
Baltimore, Maryland

John T. Behrens
Cisco Systems, Inc.
Mishawaka, Indiana

DuBois Bowman
Center for Biomedical Imaging
Statistics (CBIS)
Emory University
Atlanta, Georgia

Steve Brooks
ATASS Ltd
Exeter, United Kingdom

Brian Caffo
Department of Biostatistics
Johns Hopkins University
Baltimore, Maryland

George Casella
Department of Statistics
University of Florida
Gainesville, Florida

Radu V. Craiu
Department of Statistics
University of Toronto
Toronto, Ontario, Canada

Francesca Dominici
Department of Biostatistics
Johns Hopkins University
Baltimore, Maryland

Lynn Eberly
Division of Biostatistics
University of Minnesota
Minneapolis, Minnesota

Yanan Fan
School of Mathematics and Statistics
University of New South Wales
Sydney, Australia

Paul Fearnhead
Department of Mathematics
and Statistics
Lancaster University
Lancaster, United Kingdom

James M. Flegal
Department of Statistics
University of California
Riverside, California

Colin Fox
Department of Physics
University of Otago
Dunedin, New Zealand

Filiz Garip
Department of Sociology
Harvard University
Cambridge, Massachusetts

Andrew Gelman
Department of Statistics

and

Department of Political Science
Columbia University
New York, New York

Charles J. Geyer
School of Statistics
University of Minnesota
Minneapolis, Minnesota

Murali Haran
Center for Ecology and
Environmental Statistics
Pennsylvania State University
University Park, Pennsylvania

11

Importance Sampling, Simulated Tempering, and Umbrella Sampling

Charles J. Geyer

11.1 Importance Sampling

The importance of so-called importance sampling in Markov chain Monte Carlo (MCMC) is not what gives it that name. It is the idea that "any sample can come from any distribution" (Trotter and Tukey, 1956). Suppose that we have a Markov chain $X_1, X_2, \ldots$ having properly normalized density f for its equilibrium distribution. Let f_θ denote a parametric family of densities each absolutely continuous with respect to f. Then

$$\hat{\mu}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \frac{f_\theta(X_i)}{f(X_i)} \tag{11.1}$$

is a sensible estimator of

$$\mu(\theta) = E_\theta\{g(X)\} \tag{11.2}$$

for all θ, because by the Markov chain law of large numbers (Meyn and Tweedie, 1993, Theorem 17.1.7),

$$\hat{\mu}_n(\theta) \xrightarrow{\text{a.s.}} E_f \left\{ g(X) \frac{f_\theta(X)}{f(X)} \right\} = \int g(x) \frac{f_\theta(x)}{f(x)} f(x)\, dx = \int g(x) f_\theta(x)\, dx$$

(the requirement that f_θ is absolutely continuous with respect to f is required so that we divide by zero in the middle expressions with probability zero, so the value of the integral is not affected). With one sample from one distribution $f(x)$ we learn about $\mu(\theta)$ for all θ.

Monte Carlo standard errors (MCSEs) for importance sampling are straightforward: we just calculate the MCSE for the functional of the Markov chain (Equation 11.1) that gives our importance sampling estimator. This means we replace g in Equation 1.6 in Chapter 1 (this volume) by $g f_\theta / f$.

We are using here both the principle of "importance sampling" (in using the distribution with density f to learn about the distribution with density f_θ) and the principle of "common random numbers" (in using the same sample to learn about f_θ for all θ). The principle of common random numbers is very important. It means, for example, that

$$\nabla \tilde{\mu}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \frac{\nabla f_\theta(X_i)}{f(X_i)}$$

(Green, 1995) would undoubtedly do better (see the Bayesian model selection example in Section 1.17.3). Unfortunately, there seems to be no way to code up a function like `temper` that uses "reversible jump" and requires no theoretical work from users that, if messed up, destroys the algorithm. The `temper` function is foolproof in the sense that if the log unnormalized density function written by the user (like our `ludfun`) is correct, then the ST Markov chain has the equilibrium distribution it is supposed to have. There is nothing the user can mess up except this user-written function. No analog of this for "reversible jump" chains is apparent (to your humble author).

Two issues remain, the first being about within-model priors for the "padding" components of within-model parameter vectors $g_{\text{pad}}(\theta_{\text{pad}} \mid m)$ in the notation in Equation 11.17. Rather than choose these so that they do not depend on the data (as we did), it would be better (if more trouble) to choose them differently for each "padding" component, centering $g_{\text{pad}}(\theta_{\text{pad}} \mid m)$ so the distribution of a component of θ_{pad} is near to the marginal distribution of the same component in neighboring models (according to the `neighbors` argument of the `temper` function).

The other remaining issue is adjusting acceptance rates for jumps. There is no way to adjust this other than by changing the number of models and their definitions. But the models we have cannot be changed; if we are to calculate Bayes factors for them, then we must sample them as they are. But we can insert new models between old models. For example, if the acceptance for swaps between model i and model j is too low, then we can insert distribution k between them that has unnormalized density

$$h_k(x) = \sqrt{h_i(x)h_j(x)}.$$

This idea is inherited from simulated tempering; Geyer and Thompson (1995) have much discussion of how to insert additional distributions into a tempering network. It is another key issue in using tempering to speed up sampling. It is less obvious in the Bayes factor context, but still an available technique if needed.

Acknowledgments

This chapter benefited from detailed comments by Christina Knudson, Leif Johnson, Galin Jones, and Brian Shea.

References

Carlin, B. P. and Chib, S. 1995. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 56:261–274.

Earl, D. J. and Deem, M. W. 2005. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7:3910–3916.

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.

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 for dependent data. In E. M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 156–163. American Statistical Association, New York.

Geyer, C. J. 1994. On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society*, *Series B*, 56:261–274.

Geyer, C. J. 1999. Likelihood inference for spatial point processes. In O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M.van Lieshout (eds), *Stochastic Geometry: Likelihood and Computation*, pp. 79–140. Chapman & Hall/CRC, Boca Raton, FL.

Geyer, C. J. 2010. *mcmc: Markov Chain Monte Carlo*. R package version 0.8, available from CRAN.

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

Geyer, C. J. and Thompson, E. A. 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90:909–920.

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

Insua, D. R. and Ruggeri, F. (eds) 2000. *Robust Bayesian Analysis*, Lecture Notes in Statistics, Vol. 152. Springer, New York.

Kirkpatrick, S., Gelatt, C. D. J., and Vecchi, M. P. 1983. Optimization by simulated annealing. *Science*, 220:671–680.

Marinari, E. and Parisi, G. 1992. Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters*, 19:451–458.

Meyn, S. P. and Tweedie, R. L. 1993. *Markov Chains and Stochastic Stability*. Springer, London.

Torrie, G. M. and Valleau, J. P. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23:187–199.

Trotter, H. F. and Tukey, J. W. 1956. Conditional Monte Carlo for normal samples. In H. A. Meyer (ed.), *Symposium on Monte Carlo Methods*, pp. 64–79. Wiley, New York.

Index