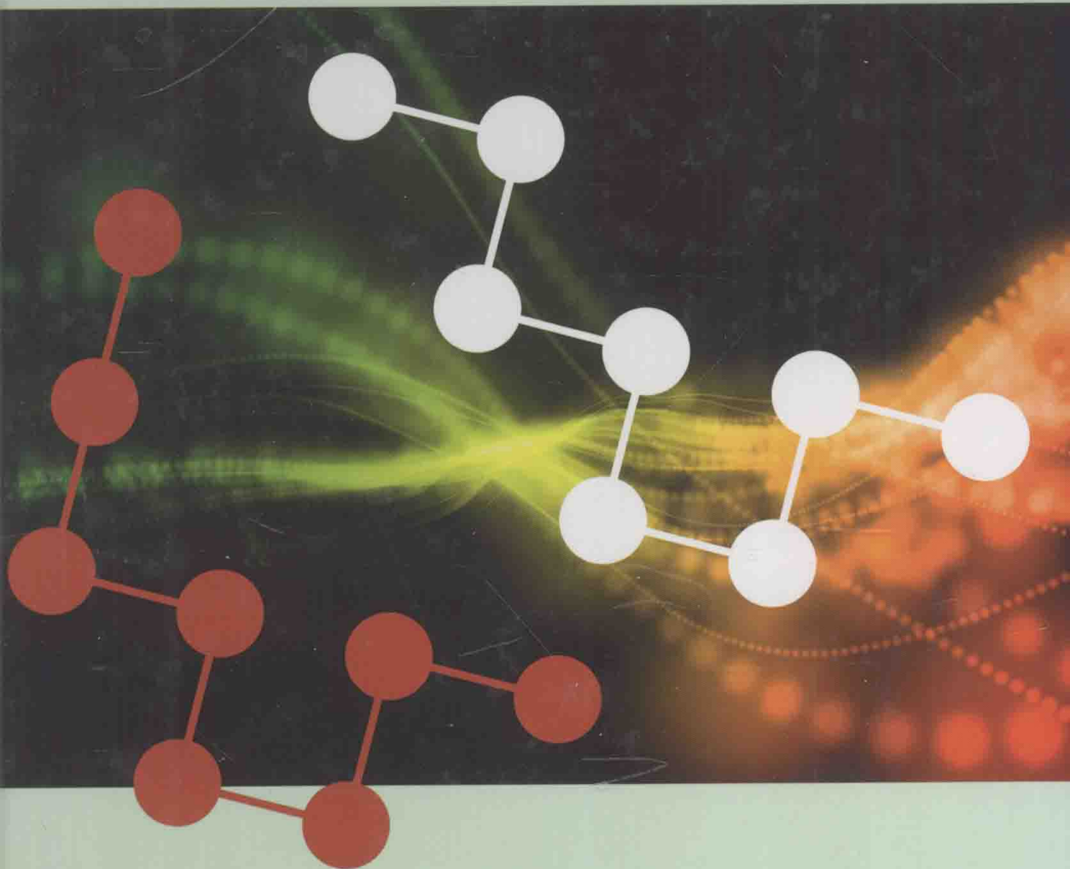
WILEY SERIES IN COMPUTATIONAL STATISTICS



Faming Liang, Chuanhai Liu, Raymond J. Carroll

ADVANCED MARKOV CHAIN MONTE CARLO METHODS

LEARNING FROM PAST SAMPLES

WILEY

Advanced Markov Chain Monte Carlo Methods

Learning from Past Samples

Faming Liang
Department of Statistics, Texas A&M University

Chuanhai Liu
Department of Statistics, Purdue University

Raymond J. Carroll
Department of Statistics, Texas A&M University

WILEY

A John Wiley and Sons, Ltd., Publication

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data
Liang, F. (Faming), 1970-
Advanced Markov Chain Monte Carlo methods : learning from past samples / Faming Liang, Chuanhai Liu, Raymond J. Carroll.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-74826-8 (cloth)
1. Monte Carlo method. 2. Markov processes. I. Liu, Chuanhai, 1959- II. Carroll, Raymond J. III. Title.
QA298.L53 2010
518′.282 – dc22

2010013148

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

ISBN 978-0-470-74826-8

Typeset in 10/12 cmr10 by Laserwords Private Limited, Chennai, India
Printed and bound in the UK by TJ International Ltd, Padstow, Cornwall

Advanced Markov Chain Monte Carlo Methods

Wiley Series in Computational Statistics

Consulting Editors:

Paolo Giudici
University of Pavia, Italy

Geof H. Givens
Colorado State University, USA

Bani K. Mallick
Texas A & M University, USA

Wiley Series in Computational Statistics is comprised of practical guides and cutting edge research books on new developments in computational statistics. It features quality authors with a strong applications focus. The texts in the series provide detailed coverage of statistical concepts, methods and case studies in areas at the interface of statistics, computing, and numerics.

With sound motivation and a wealth of practical examples, the books show in concrete terms how to select and to use appropriate ranges of statistical computing techniques in particular fields of study. Readers are assumed to have a basic understanding of introductory terminology.

The series concentrates on applications of computational methods in statistics to fields of bioinformatics, genomics, epidemiology, business, engineering, finance and applied statistics.

Titles in the Series

Billard and Diday - Symbolic Data Analysis: Conceptual Statistics and Data Mining
Bolstad - Understanding Computational Bayesian Statistics
Borgelt, Steinbrecher and Kruse - Graphical Models, 2e
Dunne - A Statistical Approach to Neutral Networks for Pattern Recognition
Liang, Liu and Carroll - Advanced Marcov Chain Monte Carlo Methods
Ntzoufras - Bayesian Modeling Using WinBUGS

To our families

Preface

The Markov Chain Monte Carlo (MCMC) method is rooted in the work of physicists such as Metropolis and von Neumann during the period 1945–55 when they employed modern electronic computers for the simulation of some probabilistic problems in atomic bomb designs. After five decades of continual development, it has become the dominant methodology in the solution of many classes of computational problems of central importance to science and technology.

Suppose that one is interested in simulating from a distribution with the density/mass function given by $f(x) \propto \exp\{-H(x)/t\}$, $x \in \mathcal{X}$, where $H(x)$ is called the energy function, and t is called the temperature. The Metropolis algorithm (Metropolis *et al.*, 1953) is perhaps the first sampling algorithm for iterative simulations. It has an extremely simple form. Starting with any point $x_0 \in \mathcal{X}$, it proceeds by iterating between the following two steps, what we call the proposal-and-decision steps:

1. (*Proposal*) Propose a random 'unbiased perturbation' of the current state x_t generated from a symmetric proposal distribution $T(x_t, y)$, i.e., $T(x_t, y) = T(y, x_t)$.

2. (*Decision*) Calculate the energy difference $\Delta H = H(y) - H(x_t)$. Set $x_{t+1} = y$ with probability $\min\{1, \exp(-\Delta H/t)\}$, and set $x_{t+1} = x_t$ with the remaining probability.

This algorithm was later generalized by Hastings (1970) to allow asymmetric proposal distributions to be used in generating the new state y. The generalized algorithm is usually called the Metropolis-Hastings algorithm. A fundamental feature of the Metropolis-Hastings update is its localness, the new state being generated in a neighborhood of the current state. This feature allows one to break a complex task into a series of manageable pieces. On the other hand, however, it tends to suffer from the local-trap problem when the energy function has multiple local minima separated by high energy barriers. In this situation, the Markov chain will be indefinitely trapped in local energy minima. Consequently, the simulation process may fail to sample from the relevant parts of the sample space, and the quantities of interest cannot be estimated with satisfactory accuracies. Many applications of the MCMC method, such as protein folding, combinatorial optimization, and spin-glasses, can be dramatically enhanced by sampling algorithms which allow the process to avoid being trapped in local energy minima.

Developing MCMC sampling algorithms that are immune to the local-trap problem has long been considered as one of the most important topics in MCMC research. During the past two decades, various advanced MCMC algorithms which address this problem have been developed. These include: the Swendsen-Wang algorithm (1987); parallel tempering (Geyer, 1991; Hukushima and Nemoto, 1996), multicanonical Monte Carlo (Berg and Neuhaus, 1991, 1992); simulated tempering (Marinari and Parisi, 1992; Geyer and Thompson, 1995); dynamic weighting (Wong and Liang, 1997; Liu *et al.*, 2001; Liang, 2002b); slice sampler (Higdon, 1998; Edwards and Sokal, 1988); evolutionary Monte Carlo (Liang and Wong, 2000, 2001b), adaptive Metropolis algorithm (Haario *et al.*, 2001); the Wang-Landau algorithm (Wang and Landau, 2001; Liang, 2005b); equi-energy sampler (Kou *et al.*, 2006); sample Metropolis-Hastings algorithm (Lewandowski and Liu, 2008); and stochastic approximation Monte Carlo (Liang *et al.*, 2007; Liang, 2009b), among others.

In addition to the local-trap problem, the Metropolis-Hastings algorithm also suffers from the inability in sampling from distributions with the mass/density function involving intractable integrals. Let $f(x) \propto c(x)\psi(x)$, where $c(x)$ denotes an intractable integral. Clearly, the Metropolis-Hastings algorithm cannot be applied to simulate from $f(x)$, as the acceptance probability would involve the intractable ratio $c(y)/c(x)$, where y denotes the candidate sample. To overcome this difficulty, advanced MCMC algorithms have also been proposed in recent literature. These include the Møller algorithm (Møller *et al.*, 2006), the exchange algorithm (Murray *et al.*, 2006), the double Metropolis-Hastings algorithm (Liang, 2009c; Jin and Liang, 2009), the Monte Carlo dynamically weighted importance sampling algorithm (Liang and Cheon, 2009), and the Monte Carlo Metropolis-Hastings sampler (Liang and Jin, 2010), among others.

One common key idea behind these advanced MCMC algorithms is *learning from past samples*. For example, stochastic approximation Monte Carlo (Liang *et al.*, 2007) modifies its invariant distribution from iteration to iteration based on its past samples in such a way that each region of the sample space can be drawn from with a desired frequency, and thus the local-trap problem can be avoided essentially. The adaptive Metropolis algorithm modifies its proposal distribution from iteration to iteration based on its past samples such that an "optimal" proposal distribution can be achieved dynamically. In the dynamic weighting algorithm, the importance weight carries the information of past samples, which helps the system move across steep energy barriers even in the presence of multiple local energy minima. In parallel tempering and evolutionary Monte Carlo, the state of the Markov chain is extended to a population of independent samples, for which, at each iteration, each sample can be updated based on entire samples of the current population. Hence, parallel tempering and evolutionary Monte Carlo can also be viewed as algorithms for learning from past samples, although they can only learn within a fixed horizon.

Meanwhile, many advanced techniques have been developed in the literature to accelerate the convergence of the Metropolis-Hastings algorithm and the Gibbs sampler; the latter can be viewed as a special form of the Metropolis-Hastings algorithm, with each component of the state vector being updated via a conditional sampling step. Such techniques include: blocking and collapsing (Liu *et al.*, 1994); reparameterization (Gelfand *et al.*, 1995); parameter expansion (Meng and van Dyk, 1999; Liu and Wu, 1999); multiple-try (Liu *et al.*, 2000); and alternating subspace-spanning resampling (Liu, 2003), among others.

The aim of this book is to provide a unified and up-to-date treatment of advanced MCMC algorithms and their variants. According to their main features, we group these advanced MCMC algorithms into several categories. The Gibbs sampler and acceleration methods, the Metropolis-Hastings algorithm and extensions, auxiliary variable-based MCMC algorithms, population-based MCMC algorithms, dynamic weighting, stochastic approximation Monte Carlo, and MCMC algorithms with adaptive proposals are described in Chapters 2–8. Chapter 1 is dedicated to brief descriptions of Bayesian inference, random number generation, and basic MCMC theory. Importance sampling, which represents another important area of Monte Carlo other than MCMC, is not fully addressed in this book. Those interested in this area should refer to Liu (2001) or Robert and Casella (2004).

This book is intended to serve three audiences: researchers specializing in Monte Carlo algorithms; scientists interested in using Monte Carlo methods; and graduate students in statistics, computational biology, engineering, and computer sciences who want to learn Monte Carlo methods. The prerequisites for understanding most of the material presented are minimal: a one-semester course on probability theory (Ross, 1998) and a one-semester course on statistical inference (Rice, 2007), both at undergraduate level. However, it would also be more desirable for readers to have a background in some specific scientific area such as Bayesian computation, artificial intelligence, or computational biology. This book is suitable as a textbook for one-semester courses on Monte Carlo methods, offered at the advanced Master's or Ph.D. level.

Faming Liang, Chuanhai Liu, and Raymond J. Carroll
December, 2009
`www.wiley.com/go/markov`

Acknowledgments

Faming Liang is most grateful to his PhD advisor professor, Wing Hung Wong, for his overwhelming passion for Markov Chain Monte Carlo and scientific problems, and for his constant encouragement. Liang's research was partially supported by grants from the National Science Foundation (DMS-0607755 and CMMI-0926803).

Chuanhai Liu's interest in computational statistics is due largely to the support and encouragement from his M.S. advisor professor, Yaoting Zhang and PhD advisor professor, Donald B. Rubin. In the mid-1980s, Chuanhai Liu learned from Professor Yaoting Zhang the importance of statistical computing. Over a time period of more than ten years from late 1980s, Chuanhai Liu learned from Professor Donald B. Rubin statistical thinking in developing iterative methods, such as EM and Gibbs-type algorithms.

Raymond Carroll's research was partially supported by a grant from the National Cancer Institute (CA57030).

Finally, we wish to thank our families for their constant love, understanding and support. It is to them that we dedicate this book.

F.L., C.L. and R.C.

Publisher's Acknowledgments

The publisher wishes to thank the following for permission to reproduce copyright material:

Table 3.2, Figure 3.4: Reproduced by permission of Licensee BioMed Central Ltd.

Table 4.1, Figure 4.2, Table 4.3: Reproduced by permission of Taylor & Francis.

Figure 5.1, Figure 5.5, Figure 5.6: Reproduced by permission of International Chinese Statistical Association.

Figure 5.2, Figure 5.3, Figure 6.5, Figure 7.1, Figure 7.3, Figure 7.10: Reproduced by permission of American Statistical Association.

Table 5.4, Figure 5.4: Reproduced by permission of American Physical Society.

Figure 5.7, Table 5.5, Figure 5.8: Reproduced by permission of American Institute of Physics.

Figure 5.9, Figure 7.11, Table 7.7, Figure 8.1, Figure 8.2, Figure 8.3: Reproduced by permission of Springer.

Figure 6.1, Figure 6.2, Table 7.2, Figure 7.2, Table 7.4, Figure 7.6, Table 7.5, Figure 7.7, Figure 7.8, Figure 7.9: Reproduced by permission of Elsevier.

Figure 6.3: Reproduced by permission of National Academy of Science.

Figure 7.4, Figure 7.5: Reproduced by permission of National Cancer Institute.

Every effort has been made to trace rights holders, but if any have been inadvertently overlooked the publishers would be pleased to make the necessary arrangements at the first opportunity.

Contents

Chapter 1

Bayesian Inference and Markov Chain Monte Carlo

1.1 Bayes

Bayesian inference is a probabilistic inferential method. In the last two decades, it has become more popular than ever due to affordable computing power and recent advances in Markov chain Monte Carlo (MCMC) methods for approximating high dimensional integrals.

Bayesian inference can be traced back to Thomas Bayes (1764), who derived the inverse probability of the success probability θ in a sequence of independent Bernoulli trials, where θ was taken from the uniform distribution on the unit interval $(0, 1)$ but treated as unobserved. For later reference, we describe his experiment using familiar modern terminology as follows.

■ **Example 1.1 The Bernoulli (or Binomial) Model With Known Prior**

Suppose that $\theta \sim \text{Unif}(0, 1)$, the uniform distribution over the unit interval $(0, 1)$, and that $x_1, \ldots, x_n$ is a sample from Bernoulli(θ), which has the sample space $\mathcal{X} = \{0, 1\}$ and probability mass function (pmf)

$$\Pr(X = 1|\theta) = \theta \quad \text{and} \quad \Pr(X = 0|\theta) = 1 - \theta, \tag{1.1}$$

where X denotes the Bernoulli random variable (r.v.) with $X = 1$ for *success* and $X = 0$ for *failure*. Write $N = \sum_{i=1}^{n} x_i$, the observed number of successes in the n Bernoulli trials. Then $N|\theta \sim \text{Binomial}(n, \theta)$, the Binomial distribution with parameters size n and probability of success θ.

Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples
Faming Liang, Chuanhai Liu and Raymond J. Carroll