Springer Texts in Statistics

Advisors:

George Casella Stephen Fienberg Ingram Olkin

Springer Texts in Statistics

- Alfred*: Elements of Statistics for the Life and Social Sciences
Berger: An Introduction to Probability and Stochastic Processes
Bilodeau and Brenner: Theory of Multivariate Statistics
Blom: Probability and Statistics: Theory and Applications
Brockwell and Davis: Introduction to Times Series and Forecasting, Second Edition
Carmona: Statistical Analysis of Financial Data in S-Plus
Chow and Teicher: Probability Theory: Independence, Interchangeability, Martingales, Third Edition
Christensen: Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization, Second Edition
Christensen: Log-Linear Models and Logistic Regression, Second Edition
Christensen: Plane Answers to Complex Questions: The Theory of Linear Models, Third Edition
Creighton: A First Course in Probability Models and Statistical Inference
Davis: Statistical Methods for the Analysis of Repeated Measurements
Dean and Voss: Design and Analysis of Experiments
du Toit, Steyn, and Stumpf: Graphical Exploratory Data Analysis
Durrett: Essentials of Stochastic Processes
Edwards: Introduction to Graphical Modelling, Second Edition
Finkelstein and Levin: Statistics for Lawyers
Flury: A First Course in Multivariate Statistics
Heiberger and Holland: Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS
Jobson: Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design
Jobson: Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods
Kalbfleisch: Probability and Statistical Inference, Volume I: Probability, Second Edition
Kalbfleisch: Probability and Statistical Inference, Volume II: Statistical Inference, Second Edition
Karr: Probability
Keyfitz: Applied Mathematical Demography, Second Edition
Kiefer: Introduction to Statistical Inference
Kokoska and Nevison: Statistical Tables and Formulae
Kulkarni: Modeling, Analysis, Design, and Control of Stochastic Systems
Lange: Applied Probability
Lange: Optimization
Lehmann: Elements of Large-Sample Theory

(continued after index)

Christian P. Robert George Casella

Monte Carlo Statistical Methods

Second Edition

With 132 Illustrations

Springer

Christian P. Robert
CEREMADE
Université Paris Dauphine
75775 Paris Cedex 16
France
xian@ceremade.dauphine.fr

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA
casella@stat.ufl.edu

Editorial Board

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data

Robert, Christian P., 1961-

Monte Carlo statistical methods / Christian P. Robert, George Casella.—2nd ed.
p. cm. — (Springer texts in statistics)

Includes bibliographical references and index.

ISBN 978-1-4419-1939-7

ISBN 978-1-4757-4145-2 (eBook)

DOI 10.1007/978-1-4757-4145-2

1. Mathematical statistics. 2. Monte Carlo method. I. Casella, George. II. Title. III. Series.

QA276.R575 2004

519.5—dc22

2004049157

ISBN 978-1-4419-1939-7 Printed on acid-free paper.

© 2004 Springer Science+Business Media New York

Originally published by Springer Science+Business Media Inc. in 2004

Softcover reprint of the hardcover 2nd edition 2004

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher Springer Science+Business Media, LLC, except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

SPIN 10945971

springeronline.com

In memory of our colleague and dear friend, Costas Goutis (1962-1996)

To Benjamin, Joachim, Rachel and Sarah, our favorite random generators!

Preface to the Second Edition

‘What do you plan to do next?’ de Wetherset asked, picking up a piece of vellum covered with minute writing and studying it. Bartholomew rose to leave. The Chancellor clearly was not interested in how they went about getting the information, only in what they discovered.

—Susanna Gregory, *An Unholy Alliance*

It is a tribute to our profession that a textbook that was current in 1999 is starting to feel old. The work for the first edition of Monte Carlo Statistical Methods (MCSM1) was finished in late 1998, and the advances made since then, as well as our level of understanding of Monte Carlo methods, have grown a great deal. Moreover, two other things have happened. Topics that just made it into MCSM1 with the briefest treatment (for example, perfect sampling) have now attained a level of importance that necessitates a much more thorough treatment. Secondly, some other methods have not withstood the test of time or, perhaps, have not yet been fully developed, and now receive a more appropriate treatment.

When we worked on MCSM1 in the mid-to-late 90s, MCMC algorithms were already heavily used, and the flow of publications on this topic was at such a high level that the picture was not only rapidly changing, but also necessarily incomplete. Thus, the process that we followed in MCSM1 was that of someone who was thrown into the ocean and was trying to grab onto the biggest and most seemingly useful objects while trying to separate the flotsam from the jetsam. Nonetheless, we also felt that the fundamentals of many of these algorithms were clear enough to be covered at the textbook level, so we swam on.

In this revision, written five years later, we have the luxury of a more relaxed perspective on the topic, given that the flurry of activity in this area has slowed somewhat into a steady state. This is not to say that there is no longer any research in this area, but simply that the tree of publications, which was growing in every direction in 1998, can now be pruned, with emphasis

being put on the major branches. For this new edition, we thus spent a good bit of time attempting to arrange the material (especially in the first ten chapters) to be presented in a coherent, flowing story, with emphasis on fundamental principles. In doing so the “Fundamental Theorem of Simulation” emerged, which we now see as a basis for many Monte Carlo algorithms (as developed in Chapters 3 and 8).

As a consequence of this coming-of-age of MCMC methods, some of the original parts of MCSM1 have therefore been expanded, and others shrunk. For example, reversible jump, sequential MC methods, two-stage Gibbs and perfect sampling now have their own chapters. Also, we now put less emphasis on some of the finer details of convergence theory, because of the simultaneous publication of Roberts and Tweedie (2004), which covers the theory of MCMC, with a comprehensive treatment of convergence results, and in general provides a deeper entry to the theory of MCMC algorithms.

We also spend less time on convergence control, because some of the methods presented in MCSM1 did not stand the test of time. The methods we preserved in Chapter 12 have been sufficiently tested to be considered reliable. Finally, we no longer have a separate chapter on missing (or latent) variables. While these models are usually a case study ideal for assessing simulation methods, they do not enjoy enough of a unified structure to be kept as a separate chapter. Instead, we dispatched most of the remaining models to different chapters of this edition.

From a (more) pedagogical point of view, we revised the book towards more accessibility and readability, thus removing the most theoretical examples and discussions. We also broke up the previously long chapters on Monte Carlo integration and Gibbs sampling into more readable chapters, with increasing coverage and difficulty. For instance, Gibbs sampling is first introduced via the slice sampler, which is simpler to describe and analyze, then the two-stage Gibbs sampler (or Data Augmentation) is presented on its own and only then do we launch into processing the general Gibbs sampler. Similarly, the experience of the previous edition led us to remove several problems or examples in every chapter, to include more detailed examples and fundamental problems, and to improve the help level on others.

Throughout the preparation of this book, and of its predecessors, we were fortunate to have colleagues who provided help. George Fishman, Anne Philippe, Judith Rousseau, as well as numerous readers, pointed out typos and mistakes in the previous version. We especially grateful to Christophe Andrieu, Roberto Casarin, Nicolas Chopin, Arnaud Doucet, Jim Hobert, Merrilee Hurn, Jean-Michel Marin, and Jesper Møller, François Perron, Arafat Tayeb, and an anonymous referee, for detailed reading of parts (or the whole) of the manuscript of this second version. Obviously, we, rather than they, should be held responsible for any imperfection remaining in the current edition! We also gained very helpful feedback from the audiences of our lectures on MCMC methods, especially during the summer schools of Luminy (France) in 2001

and 2002; Les Diablerets (Switzerland) and Venezia (Italy) in 2003; Orlando in 2002 and 2003; Atlanta in 2003; and Oulanka (Finland) in 2004.

Thanks to Elias Moreno for providing a retreat in Granada for the launching of this project in November 2002 (almost) from the top of Mulhacen; to Manuella Delbois who made the move to BibTeX possible by translating the entire reference list of MCMC1 into BibTeX format; to Jeff Gill for his patient answers to our R questions, and to Olivier Cappé for never-ending Linux and L^AT_EX support, besides his more statistical (but equally helpful!) comments and suggestions.

Christian P. Robert
George Casella

June 15, 2004

Preface to the First Edition

He sat, continuing to look down the nave, when suddenly the solution to the problem just seemed to present itself. It was so simple, so obvious he just started to laugh...

—P.C. Doherty, *Satan in St Mary's*

Monte Carlo statistical methods, particularly those based on Markov chains, have now matured to be part of the standard set of techniques used by statisticians. This book is intended to bring these techniques into the classroom, being (we hope) a self-contained logical development of the subject, with all concepts being explained in detail, and all theorems, etc. having detailed proofs. There is also an abundance of examples and problems, relating the concepts with statistical practice and enhancing primarily the application of simulation techniques to statistical problems of various difficulties.

This is a textbook intended for a second-year graduate course. We do not assume that the reader has any familiarity with Monte Carlo techniques (such as random variable generation) or with any Markov chain theory. We do assume that the reader has had a first course in statistical theory at the level of *Statistical Inference* by Casella and Berger (1990). Unfortunately, a few times throughout the book a somewhat more advanced notion is needed. We have kept these incidents to a minimum and have posted warnings when they occur. While this is a book on simulation, whose actual implementation must be processed through a computer, no requirement is made on programming skills or computing abilities: algorithms are presented in a program-like format but in plain text rather than in a specific programming language. (Most of the examples in the book were actually implemented in C, with the S-Plus graphical interface.)

Chapters 1–3 are introductory. Chapter 1 is a review of various statistical methodologies and of corresponding computational problems. Chapters 2 and 3 contain the basics of random variable generation and Monte Carlo integration. Chapter 4, which is certainly the most theoretical in the book, is an

introduction to Markov chains, covering enough theory to allow the reader to understand the workings and evaluate the performance of Markov chain Monte Carlo (MCMC) methods. Section 4.1 is provided for the reader who already is familiar with Markov chains, but needs a refresher, especially in the application of Markov chain theory to Monte Carlo calculations. Chapter 5 covers optimization and provides the first application of Markov chains to simulation methods. Chapters 6 and 7 cover the heart of MCMC methodology, the Metropolis–Hastings algorithm and the Gibbs sampler. Finally, Chapter 8 presents the state-of-the-art methods for monitoring convergence of the MCMC methods and Chapter 9 shows how these methods apply to some statistical settings which cannot be processed otherwise, namely the missing data models.

Each chapter concludes with a section of notes that serve to enhance the discussion in the chapters, describe alternate or more advanced methods, and point the reader to further work that has been done, as well as to current research trends in the area. The level and rigor of the notes are variable, with some of the material being advanced.

The book can be used at several levels and can be presented in several ways. For example, Chapters 1–3 and most of Chapter 5 cover standard simulation theory, and hence serve as a basic introduction to this topic. Chapters 6–9 are totally concerned with MCMC methodology. A one-semester course, assuming no familiarity with random variable generation or Markov chain theory could be based on Chapters 1–7, with some illustrations from Chapters 8 and 9. For instance, after a quick introduction with examples from Chapter 1 or Section 3.1, and a description of Accept–Reject techniques of Section 2.3, the course could cover Monte Carlo integration (Section 3.2, Section 3.3 [except Section 3.3.3], Section 3.4, Section 3.7), Markov chain theory through either Section 4.1 or Section 4.2–Section 4.8 (while adapting the depth to the mathematical level of the audience), mention stochastic optimization via Section 5.3, and describe Metropolis–Hastings and Gibbs algorithms as in Chapters 6 and 7 (except Section 6.5, Section 7.1.5, and Section 7.2.4). Depending on the time left, the course could conclude with some diagnostic methods of Chapter 8 (for instance, those implemented in CODA) and/or some models of Chapter 9 (for instance, the mixture models of Section 9.3 and Section 9.4). Alternatively, a more advanced audience could cover Chapter 4 and Chapters 6–9 in one semester and have a thorough introduction to MCMC theory and methods.

Much of the material in this book had its original incarnation as the French monograph *Méthodes de Monte Carlo par Chaînes de Markov* by Christian Robert (Paris: Economica, 1996), which has been tested for several years on graduate audiences (in France, Québec, and even Norway). Nonetheless, it constitutes a major revision of the French text, with the inclusion of problems, notes, and the updating of current techniques, to keep up with the advances that took place in the past two years (like Langevin diffusions, perfect sampling, and various types of monitoring).

Throughout the preparation of this book, and its predecessor, we were fortunate to have colleagues who provided help. Sometimes this was in the form of conversations or references (thanks to Steve Brooks and Sid Chib!), and a few people actually agreed to read through the manuscript. Our colleague and friend, Costas Goutis, provided many helpful comments and criticisms, mostly on the French version, but these are still felt in this version. We are also grateful to Brad Carlin, Dan Fink, Jim Hobert, Galin Jones and Krishanu Maulik for detailed reading of parts of the manuscript, to our historian Walter Piegorsch, and to Richard Tweedie, who taught from the manuscript and provided many helpful suggestions, and to his students, Nicole Benton, Sarah Streett, Sue Taylor, Sandy Thompson, and Alex Trindade. Richard, whose influence on the field was considerable, both from a theoretical and a methodological point of view, most sadly passed away last July. His spirit, his humor and his brightness will remain with us for ever, as a recurrent process. Christophe Andrieu, Virginie Braïdo, Jean-Jacques Colleau, Randall Douc, Arnaud Doucet, George Fishman, Jean-Louis Foulley, Arthur Gretton, Ana Justel, Anne Philippe, Sandrine Micaleff, and Judith Rousseau pointed out typos and mistakes in either the French or the English versions (or both), but should not be held responsible for those remaining! Part of Chapter 8 has a lot in common with a “reviewww” written by Christian Robert with Chantal Guihenneuc-Jouyaux and Kerrie Mengersen for the Valencia Bayesian meeting (and the Internet!). The input of the French working group “MC Cube,” whose focus is on convergence diagnostics, can also be felt in several places of this book. Wally Gilks and David Spiegelhalter granted us permission to use their graph (Figure 2.5) and examples as Problems 10.29–10.36, for which we are grateful. Agostino Nobile kindly provided the data on which Figures 10.4 and 10.4 are based. Finally, Arnaldo Frigessi (from Roma) made the daring move of teaching (in English) from the French version in Oslo, Norway; not only providing us with very helpful feedback but also contributing to making Europe more of a reality!

Christian P. Robert
George Casella

January 2002

Contents

Preface to the Second Edition	IX
Preface to the First Edition	XIII
1 Introduction	1
1.1 Statistical Models	1
1.2 Likelihood Methods	5
1.3 Bayesian Methods	12
1.4 Deterministic Numerical Methods	19
1.4.1 Optimization	19
1.4.2 Integration	21
1.4.3 Comparison	21
1.5 Problems	23
1.6 Notes	30
1.6.1 Prior Distributions	30
1.6.2 Bootstrap Methods	32
2 Random Variable Generation	35
2.1 Introduction	35
2.1.1 Uniform Simulation	36
2.1.2 The Inverse Transform	38
2.1.3 Alternatives	40
2.1.4 Optimal Algorithms	41
2.2 General Transformation Methods	42
2.3 Accept–Reject Methods	47
2.3.1 The Fundamental Theorem of Simulation	47
2.3.2 The Accept–Reject Algorithm	51
2.4 Envelope Accept–Reject Methods	53
2.4.1 The Squeeze Principle	53
2.4.2 Log-Concave Densities	56
2.5 Problems	62

2.6	Notes	72
2.6.1	The Kiss Generator	72
2.6.2	Quasi-Monte Carlo Methods	75
2.6.3	Mixture Representations	77
3	Monte Carlo Integration	79
3.1	Introduction	79
3.2	Classical Monte Carlo Integration	83
3.3	Importance Sampling	90
3.3.1	Principles	90
3.3.2	Finite Variance Estimators	94
3.3.3	Comparing Importance Sampling with Accept–Reject	103
3.4	Laplace Approximations	107
3.5	Problems	110
3.6	Notes	119
3.6.1	Large Deviations Techniques	119
3.6.2	The Saddlepoint Approximation	120
4	Controlling Monte Carlo Variance	123
4.1	Monitoring Variation with the CLT	123
4.1.1	Univariate Monitoring	124
4.1.2	Multivariate Monitoring	128
4.2	Rao–Blackwellization	130
4.3	Riemann Approximations	134
4.4	Acceleration Methods	140
4.4.1	Antithetic Variables	140
4.4.2	Control Variates	145
4.5	Problems	147
4.6	Notes	153
4.6.1	Monitoring Importance Sampling Convergence	153
4.6.2	Accept–Reject with Loose Bounds	154
4.6.3	Partitioning	155
5	Monte Carlo Optimization	157
5.1	Introduction	157
5.2	Stochastic Exploration	159
5.2.1	A Basic Solution	159
5.2.2	Gradient Methods	162
5.2.3	Simulated Annealing	163
5.2.4	Prior Feedback	169
5.3	Stochastic Approximation	174
5.3.1	Missing Data Models and Demarginalization	174
5.3.2	The EM Algorithm	176
5.3.3	Monte Carlo EM	183
5.3.4	EM Standard Errors	186

5.4	Problems	188
5.5	Notes	200
5.5.1	Variations on EM	200
5.5.2	Neural Networks	201
5.5.3	The Robbins–Monro procedure	201
5.5.4	Monte Carlo Approximation	203
6	Markov Chains	205
6.1	Essentials for MCMC	206
6.2	Basic Notions	208
6.3	Irreducibility, Atoms, and Small Sets	213
6.3.1	Irreducibility	213
6.3.2	Atoms and Small Sets	214
6.3.3	Cycles and Aperiodicity	217
6.4	Transience and Recurrence	218
6.4.1	Classification of Irreducible Chains	218
6.4.2	Criteria for Recurrence	221
6.4.3	Harris Recurrence	221
6.5	Invariant Measures	223
6.5.1	Stationary Chains	223
6.5.2	Kac’s Theorem	224
6.5.3	Reversibility and the Detailed Balance Condition	229
6.6	Ergodicity and Convergence	231
6.6.1	Ergodicity	231
6.6.2	Geometric Convergence	236
6.6.3	Uniform Ergodicity	237
6.7	Limit Theorems	238
6.7.1	Ergodic Theorems	240
6.7.2	Central Limit Theorems	242
6.8	Problems	247
6.9	Notes	258
6.9.1	Drift Conditions	258
6.9.2	Eaton’s Admissibility Condition	262
6.9.3	Alternative Convergence Conditions	263
6.9.4	Mixing Conditions and Central Limit Theorems	263
6.9.5	Covariance in Markov Chains	265
7	The Metropolis–Hastings Algorithm	267
7.1	The MCMC Principle	267
7.2	Monte Carlo Methods Based on Markov Chains	269
7.3	The Metropolis–Hastings algorithm	270
7.3.1	Definition	270
7.3.2	Convergence Properties	272
7.4	The Independent Metropolis–Hastings Algorithm	276
7.4.1	Fixed Proposals	276

7.4.2	A Metropolis–Hastings Version of ARS	285
7.5	Random Walks	287
7.6	Optimization and Control	292
7.6.1	Optimizing the Acceptance Rate	292
7.6.2	Conditioning and Accelerations	295
7.6.3	Adaptive Schemes	299
7.7	Problems	302
7.8	Notes	313
7.8.1	Background of the Metropolis Algorithm	313
7.8.2	Geometric Convergence of Metropolis–Hastings Algorithms	315
7.8.3	A Reinterpretation of Simulated Annealing	315
7.8.4	Reference Acceptance Rates	316
7.8.5	Langevin Algorithms	318
8	The Slice Sampler	321
8.1	Another Look at the Fundamental Theorem	321
8.2	The General Slice Sampler	326
8.3	Convergence Properties of the Slice Sampler	329
8.4	Problems	333
8.5	Notes	335
8.5.1	Dealing with Difficult Slices	335
9	The Two-Stage Gibbs Sampler	337
9.1	A General Class of Two-Stage Algorithms	337
9.1.1	From Slice Sampling to Gibbs Sampling	337
9.1.2	Definition	339
9.1.3	Back to the Slice Sampler	343
9.1.4	The Hammersley–Clifford Theorem	343
9.2	Fundamental Properties	344
9.2.1	Probabilistic Structures	344
9.2.2	Reversible and Interleaving Chains	349
9.2.3	The Duality Principle	351
9.3	Monotone Covariance and Rao–Blackwellization	354
9.4	The EM–Gibbs Connection	357
9.5	Transition	360
9.6	Problems	360
9.7	Notes	366
9.7.1	Inference for Mixtures	366
9.7.2	ARCH Models	368
10	The Multi-Stage Gibbs Sampler	371
10.1	Basic Derivations	371
10.1.1	Definition	371
10.1.2	Completion	373

10.1.3	The General Hammersley–Clifford Theorem	376
10.2	Theoretical Justifications	378
10.2.1	Markov Properties of the Gibbs Sampler	378
10.2.2	Gibbs Sampling as Metropolis–Hastings	381
10.2.3	Hierarchical Structures	383
10.3	Hybrid Gibbs Samplers	387
10.3.1	Comparison with Metropolis–Hastings Algorithms	387
10.3.2	Mixtures and Cycles	388
10.3.3	Metropolizing the Gibbs Sampler	392
10.4	Statistical Considerations	396
10.4.1	Reparameterization	396
10.4.2	Rao–Blackwellization	402
10.4.3	Improper Priors	403
10.5	Problems	407
10.6	Notes	419
10.6.1	A Bit of Background	419
10.6.2	The BUGS Software	420
10.6.3	Nonparametric Mixtures	420
10.6.4	Graphical Models	422
11	Variable Dimension Models and Reversible Jump Algorithms	425
11.1	Variable Dimension Models	425
11.1.1	Bayesian Model Choice	426
11.1.2	Difficulties in Model Choice	427
11.2	Reversible Jump Algorithms	429
11.2.1	Green’s Algorithm	429
11.2.2	A Fixed Dimension Reassessment	432
11.2.3	The Practice of Reversible Jump MCMC	433
11.3	Alternatives to Reversible Jump MCMC	444
11.3.1	Saturation	444
11.3.2	Continuous-Time Jump Processes	446
11.4	Problems	449
11.5	Notes	458
11.5.1	Occam’s Razor	458
12	Diagnosing Convergence	459
12.1	Stopping the Chain	459
12.1.1	Convergence Criteria	461
12.1.2	Multiple Chains	464
12.1.3	Monitoring Reconsidered	465
12.2	Monitoring Convergence to the Stationary Distribution	465
12.2.1	A First Illustration	465
12.2.2	Nonparametric Tests of Stationarity	466
12.2.3	Renewal Methods	470

12.2.4	Missing Mass	474
12.2.5	Distance Evaluations	478
12.3	Monitoring Convergence of Averages	480
12.3.1	A First Illustration	480
12.3.2	Multiple Estimates	483
12.3.3	Renewal Theory	490
12.3.4	Within and Between Variances	497
12.3.5	Effective Sample Size	499
12.4	Simultaneous Monitoring	500
12.4.1	Binary Control	500
12.4.2	Valid Discretization	503
12.5	Problems	504
12.6	Notes	508
12.6.1	Spectral Analysis	508
12.6.2	The CODA Software	509
13	Perfect Sampling	511
13.1	Introduction	511
13.2	Coupling from the Past	513
13.2.1	Random Mappings and Coupling	513
13.2.2	Propp and Wilson's Algorithm	516
13.2.3	Monotonicity and Envelopes	518
13.2.4	Continuous States Spaces	523
13.2.5	Perfect Slice Sampling	526
13.2.6	Perfect Sampling via Automatic Coupling	530
13.3	Forward Coupling	532
13.4	Perfect Sampling in Practice	535
13.5	Problems	536
13.6	Notes	539
13.6.1	History	539
13.6.2	Perfect Sampling and Tempering	540
14	Iterated and Sequential Importance Sampling	545
14.1	Introduction	545
14.2	Generalized Importance Sampling	546
14.3	Particle Systems	547
14.3.1	Sequential Monte Carlo	547
14.3.2	Hidden Markov Models	549
14.3.3	Weight Degeneracy	551
14.3.4	Particle Filters	552
14.3.5	Sampling Strategies	554
14.3.6	Fighting the Degeneracy	556
14.3.7	Convergence of Particle Systems	558
14.4	Population Monte Carlo	559
14.4.1	Sample Simulation	560

14.4.2	General Iterative Importance Sampling	560
14.4.3	Population Monte Carlo	562
14.4.4	An Illustration for the Mixture Model	563
14.4.5	Adaptativity in Sequential Algorithms	565
14.5	Problems	570
14.6	Notes	577
14.6.1	A Brief History of Particle Systems	577
14.6.2	Dynamic Importance Sampling	577
14.6.3	Hidden Markov Models	579
A	Probability Distributions	581
B	Notation	585
B.1	Mathematical	585
B.2	Probability	586
B.3	Distributions	586
B.4	Markov Chains	587
B.5	Statistics	588
B.6	Algorithms	588
	References	591
	Index of Names	623
	Index of Subjects	631

List of Tables

1.1	Challenger data	15
1.2	Some conjugate families	31
2.1	Horse Kick Data	61
3.1	Evaluation of some normal quantiles	85
3.2	Surgery vs. radiation data	86
3.3	Cutoff points for contingency table data	88
3.4	Comparison of instrumental distributions	102
3.5	Comparison between importance sampling estimators	108
3.6	Laplace approximation of a Gamma integral	110
3.7	Saddlepoint approximation of a noncentral χ^2 integral	122
5.1	Stochastic gradient runs	163
5.2	Simulated annealing runs	169
5.3	Sequence of Bayes estimators for the Gamma distribution	172
5.4	Average grades of first-year students	173
5.5	Maximum likelihood estimates of mean grades	173
5.6	Cellular phone data	180
5.7	Tree swallow movements	197
5.8	Selected batting average data	198
7.1	Monte Carlo saddlepoint approximation of a noncentral chi squared integral	284
7.2	Approximation of normal moments by a random walk Metropolis–Hastings algorithm	288
7.3	Approximation of the inverse Gaussian distribution by the Metropolis–Hastings algorithm	294
7.4	Improvement in quadratic risk from Rao–Blackwellization (1)	297
7.5	Improvement in quadratic risk from Rao–Blackwellization (2)	299
7.6	Braking distances	307

XXVI List of Tables

7.7	Performance of the Metropolis–Hastings algorithm [A.29]	317
9.1	Frequencies of passage for 360 consecutive observations.	346
9.2	Observations of $\mathcal{N}_2(0, \Sigma)$ with missing data	364
9.3	Estimation result for the factor ARCH model	369
10.1	Failures of pumps in a nuclear plant	386
10.2	Interquantile ranges for the Gibbs sampling and the modification of Liu	396
10.3	Occurrences of clinical mastitis in dairy herds	409
10.4	Calcium concentration in turnip greens	414
11.1	Galaxy data	451
11.2	Yearly number of mining accidents in England	455
12.1	Estimation of the asymptotic variance for renewal control (1) . .	495
12.2	Estimation of the asymptotic variance for renewal control (2) . .	496
12.3	Gibbs approximations of expectations and variances estimated by renewal	497
12.4	Evolution of initializing and convergence times	502

List of Figures

1.1	Cauchy likelihood	11
1.2	Challenger failure probabilities	16
1.3	Newton–Raphson Algorithm	20
2.1	Plot of chaotic pairs	38
2.2	Probability of acceptance for Jöhnk’s algorithm	45
2.3	Generation of Beta random variables.....	49
2.4	Envelope for generic Accept–Reject	50
2.5	Lower and upper envelopes of a log-concave density	57
2.6	Posterior distributions of capture log-odds ratios	59
2.7	ARS approximation to capture log-odds posterior distribution	60
2.8	ARS approximation to the intercept distribution	61
2.9	Representation of $y = 69069x \bmod 1$ by uniform sampling ..	74
2.10	Plots of pairs from the Kiss generator	76
3.1	One-dimensional Monte Carlo integration	84
3.2	Contingency table test	88
3.3	Empirical cdf’s of log-likelihoods	89
3.4	Approximated error risks	93
3.5	Range of three estimators of $\mathbb{E}_f[X/(1 - X) ^{1/2}]$	97
3.6	Range of an importance sampling estimator of $\mathbb{E}_f[X/(1 - X) ^{1/2}]$	98
3.7	Convergence of four estimators of $\mathbb{E}_f[X^5 \mathbb{I}_{X \geq 2.1}]$	99
3.8	Convergence of four estimators of $\mathbb{E}_f[h_3(X)]$	100
3.9	Convergence of four estimators of $\mathbb{E}_f[h_5(X)]$	102
3.10	Convergence of estimators of $\mathbb{E}[X/(1 + X)]$	106
3.11	Convergence of the estimators of $\mathbb{E}[X/(1 + X)]$	107
4.1	Range and confidence bands for a simple example	125

XXVIII List of Figures

4.2	Range and confidence bands for the Cauchy-Normal problem (1)	127
4.3	Range and confidence bands for the Cauchy-Normal problem (2)	128
4.4	Confidence bands for running means	130
4.5	Convergence of estimators of $\mathbb{E}[\exp(-X^2)]$	132
4.6	Comparisons of an Accept-Reject and importance sampling estimator	134
4.7	Convergence of the estimators of $\mathbb{E}[X \log(X)]$	138
4.8	Convergence of estimators of $\mathbb{E}_\nu[(1 + e^X)\mathbb{I}_{X \leq 0}]$	139
4.9	Convergence of estimators of $\mathbb{E}_\nu[(1 + e^X)\mathbb{I}_{X \leq 0}]$	139
4.10	Approximate risks of truncated James-Stein estimators	142
4.11	Comparison of an antithetic and standard iid estimate	144
4.12	Graphs of the variance coefficients	155
5.1	Simple Monte Carlo maximization	160
5.2	Representation of the function of Example 5.3	161
5.3	Stochastic gradient paths	164
5.4	One-dimensional simulated annealing maximization	166
5.5	Simulated annealing sequence	170
5.6	Censored data likelihood	175
5.7	EM sequence for cellular phone data	181
5.8	Multiple mode EM	182
5.9	EM estimate and standard deviation	187
6.1	Trajectories of AR(1)	228
6.2	Means of AR(1)	245
6.3	Convergence of AR(1)	245
7.1	Convergence of Accept-Reject and Metropolis-Hastings (1)	280
7.2	Convergence of Accept-Reject and Metropolis-Hastings (2)	281
7.3	Estimation of logistic parameters	283
7.4	Histograms of three samples from [A.29]	289
7.5	Confidence envelopes for the random walk Metropolis-Hastings algorithm	290
7.6	Invalid adaptive MCMC (1)	301
7.7	Invalid adaptive MCMC (2)	301
7.8	Nonparametric invalid adaptive MCMC	302
7.9	Convergence of Langevin and iid simulations	320
8.1	Simple slice sampler	324
8.2	Ten steps of the slice sampler	325
8.3	Three slice samples for the truncated normal	325
8.4	A 3D slice sampler	327
8.5	A poor slice sampler	333

9.1	Gibbs output for mixture posterior	343
9.2	Evolution of the estimator δ_{rb} of [A.35]	347
9.3	Gibbs output for cellular phone data	359
10.1	Nonconnected support	380
10.2	Gibbs output for mastitis data	384
10.3	Successive moves of a Gibbs chain	389
10.4	Gibbs chain for the probit model	392
10.5	Hybrid chain for the probit model	392
10.6	Comparison of Gibbs sampling and Liu's modification	395
10.7	Evolution of the estimation of a mixture distribution	400
10.8	Convergence of parameter estimators for a mixture distribution	401
10.9	Iterations of a divergent Gibbs sampling algorithm	405
10.10	Evolution of $(\beta^{(t)})$ for a random effects model	407
11.1	Speed of Corona Borealis galaxies	426
11.2	Linear regression with reversible jumps	435
11.3	Reversible jump MCMC output for a mixture	440
11.4	Conditional reversible jump MCMC output for a mixture	441
11.5	Fit by an averaged density	442
11.6	Reversible jump output for the AR(p) model	443
12.1	Probit log-posterior	460
12.2	Probit β sequence (1)	467
12.3	Probit β sequence (2)	467
12.4	Probit β sequence (3)	467
12.5	Probit β sequence (4)	468
12.6	Probit posterior sequence	468
12.7	Probit ratio β/σ sequence	469
12.8	Plot of successive Kolmogorov–Smirnov statistics	470
12.9	Plot of renewal probabilities for the pump failure data	474
12.10	Bimodal density	475
12.11	Missing mass evaluation	476
12.12	Control curves for a mixture model	477
12.13	Convergence of empirical averages for Example 12.10	481
12.14	Evolution of the D_t^1 criterion for Example 12.10	482
12.15	Evolution of CUSUMs for Example 12.1	483
12.16	Gibbs sampler stuck at a wrong mode	483
12.17	CUSUMs for the mixture posterior	484
12.18	Approximation of the density (12.17)	487
12.19	Convergence of four estimators for the density (12.17)	489
12.20	Convergence of four estimators of $\mathbb{E}[(X^{(t)})^{0.8}]$	490
12.21	Evolutions of R_T and W_T for (12.17)	498
12.22	Convergence of the mean $\hat{\varrho}_t$ for [A.30]	503

12.23	Discretization of a continuous Markov chain	504
13.1	All possible transitions for the Beta-Binomial(2,2,4) example	515
13.2	Perfect sampling for a mixture posterior distribution	520
13.3	Nine iterations of a coupled Gibbs sampler	523
13.4	Coupling from the past on the distribution (13.7)	530
13.5	Successful CFTP for an exponential mixture	532
13.6	Forward simulation with dominating process	542
14.1	Simulated target tracking output	548
14.2	Hidden Markov model	549
14.3	Simulated sample of a stochastic volatility process	550
14.4	Importance sampling target tracking reconstruction	553
14.5	Particle filter target tracking reconstruction	555
14.6	Particle filter target tracking range	556
14.7	Mixture log-posterior distribution and PMC sample	565
14.8	Adaptivity of mixture PMC algorithm	566
14.9	MCMC sample for a stochastic volatility model	567
14.10	MCMC estimate for a stochastic volatility model	568
14.11	PMC sample for a stochastic volatility model	568
14.12	PMC estimate for a stochastic volatility model	569