Springer Series in Statistics

Jun S. Liu

Monte Carlo Strategies in Scientific Computing

科学计算中的蒙特卡罗策略

Springer

世界图书出版公司
www.wpcbj.com.cn

Jun S. Liu

Monte Carlo Strategies in Scientific Computing

With 56 Figures

Springer

世界图书出版公司

书　　名：Monte Carlo Strategies in Scientific Computing
作　　者：J. S. Liu
中 译 名：科学计算中的蒙特卡罗策略
出 版 者：世界图书出版公司北京公司
印 刷 者：北京世图印刷厂
发　　行：世界图书出版公司北京公司 (北京朝内大街 137 号　100010)
联系电话：010-64015659, 64038347
电子信箱：kjsk@vip.sina.com
开　　本：24 开　　　印　张：15.5
出版年代：2005 年 6 月
书　　号：7-5062-7258-X/ O · 525
版权登记：图字:01-2005-2221
定　　价：48.00 元

Preface

An early experiment that conceives the basic idea of Monte Carlo computation is known as "Buffon's needle" (Dörrie 1965), first stated by Georges Louis Leclerc Comte de Buffon in 1777. In this well-known experiment, one throws a needle of length l onto a flat surface with a grid of parallel lines with spacing D ($D > l$). It is easy to compute that, under ideal conditions, the chance that the needle will intersect one of the lines is $2l/\pi D$. Thus, if we let p_N be the proportion of "intersects" in N throws, we can have an estimate of π as

$$\hat{\pi} = \lim_{N\to\infty} \frac{2l}{p_N D},$$

which will "converge" to π as N increases to infinity. Numerous investigators actually used this setting to estimate π. The idea of simulating random processes so as to help evaluate certain quantities of interest is now an essential part of scientific computing.

A systematic use of the Monte Carlo method for real scientific problems appeared in the early days of electronic computing (1945-55) and accompanied the development of the world's first programmable "super" computer, MANIAC (Mathematical Analyzer, Numerical Integrator and Computer), at Los Alamos during World War II. In order to make a good use of these fast computing machines, scientists (Stanislaw Ulam, John von Neumann, Nicholas Metropolis, Enrico Fermi, etc.) invented a statistical sampling-based method for solving numerical problems concerning random neutron diffusion in fissile material in atomic bomb designs and for estimating eigenvalues of the Schrödinger equation. The basic idea underlying the method was first brought up by Ulam and deliberated between him and von Neumann in a car when they drove together from Los Alamos to Lamy. Allegedly, Nick Metropolis coined the name "Monte Carlo," which played an essential role in popularizing the method.

In the early 1950s, statistical physicists (N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller) introduced a Markov-chain-based dynamic Monte Carlo method for the simulation of simple fluids. This method was later extended to cover more and more complex physical systems, including spin glass models, harmonic crystal, polymer models, etc. In the 1980s, statisticians and computer scientists developed Monte-Carlo-based methods for a wide variety of tasks such as combinatorial optimizations, nonparametric statistical inference (e.g., jackknife and bootstrap), likelihood computation with missing observations, statistical genetics analysis, Bayesian modeling and computations, and others. In the 1990s, the method began to play an important role in computational biology and was used to solve problems in sequence motif identification and the analysis of complex pedigree. Now, the list of application areas of Monte Carlo methods includes biology (Leach 1996, Karplus and Petsko 1990, Lawrence, Altschul, Boguski, Liu, Neuwald and Wootton 1993), chemistry (Alder and Wainwright 1959), computer science (Kirkpatrick, Gelatt and Vecchi 1983), economics and finance (Gouriérourx and Monfort 1997); engineering (Geman and Geman 1984), material science (Frenkel and Smit 1996), physics (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller 1953, Goodman and Sokal 1989, Marinari and Parisi 1992), statistics (Efron 1979, Gelfand and Smith 1990, Rubin 1987, Tanner and Wong 1987), and many others. Among all Monte Carlo methods, *Markov chain Monte Carlo* (MCMC) provides an enormous scope for dealing with very complicated stochastic systems and has been the central pillar in the study of macromolecules and other physical systems. Recently, the MCMC methodology has drawn much attention from statisticians because the method enables them to entertain more sophisticated and realistic statistical models.

Being attracted by the extreme flexibility and power of the Monte Carlo method, many researchers in different scientific areas have contributed to its development. However, because a substantial amount of domain-specific knowledge is required in order to understand problems in any of these fields, communications among researchers in these fields are very limited. Many efforts have been devoted to the reinvention of techniques that have been developed in other fields. It is therefore desirable to develop a relatively general framework in which scientists in every field — e.g., theoretical chemists, statistical physicists, structural biologists, statisticians, econometricians, and computer scientists — can compare their Monte Carlo techniques and learn from each other. For a large number of scientists and engineers who employ Monte Carlo simulation and related global optimization techniques (such as simulated annealing) as an essential tool in their work, there is also a need to keep up to date with recent advances in Monte Carlo methodologies and to understand the nature and connection of various proposed methods. The aim of this book is to provide a self-contained, unified, and up-to-date treatment of the Monte Carlo method.

This book is intended to serve three audiences: researchers specializing in the study of Monte Carlo algorithms; scientists who are interested in using advanced Monte Carlo techniques; and graduate students in statistics, computational biology, and computer sciences who want to learn about Monte Carlo computations. The prerequisites for understanding most of the methods described in this book are rather minimal: a one-semester course on probability theory (Pitman 1993) and a one-semester course on theoretical statistics (Rice 1994), both at the undergraduate level. However, it would be more desirable if the reader has some background in a specific scientific field such as artificial intelligence, computational biology, computer vision, engineering, or Bayesian statistics in which heavy computations are involved. This book is most suitable for a second-year graduate-level course on Monte Carlo methods, with an emphasis on their relevance to scientific and statistical research.

The author is most grateful to his mentor and friend Wing Hung Wong for his many important suggestions, his overwhelming passion for Monte Carlo and scientific problems, and his continuous encouragement. The author is also grateful to Persi Diaconis for teaching him many things including Markov chain theory, group theory, and nonparametric Bayes methods, to both Susan Holmes and Persi for numerous enlightening conversations on Markov chain Monte Carlo and other related problems, to Donald B. Rubin for insights on the missing data formulation and the Bayesian thinking, to Jonathan Goodman for helpful comments on multigrid Monte Carlo, to Yingnian Wu and Songchun Zhu for their materials on pattern simulations and thoughts on conditional sampling, to Faming Liang for his supply of many examples and figures, and to Minghui Chen and David van Dyk for helpful comments. Several former graduate students in the statistics departments of Stanford and Harvard universities — Yuguo Chen, Lingyu Chen, Chiara Sabatti, Tanya Logvinenko, Zhaohui Qin and Juni Zhang — have contributed in many ways to the development of this book. Ms. Helen Tombropoulos has provided editorial assistance to the author both for this book and for many articles published earlier. Finally, the author is greatly indebted to his wife Wei for her love and her continuous support of his research activities these years. Part of the book was written when the author was on the faculty of the Statistics Department of Stanford University. This work was also partially supported by the National Science Foundation Grants DMS-9803649 and DMS-0094613.

Cambridge, Massachusetts *Jun Liu*
March 2001

Jun S. Liu
Department of Statistics
Harvard University
Cambridge, MA 02138
jliu@hustat.harvard.edu
USA

Library of Congress Cataloging-in-Publication Data
Liu, Jun S.
Monte Carlo strategies in scientific computing / Jun S. Liu
p. cm. — (Springer series in statistics)
Includes bibliographical references and index.
ISBN 0-387-95230-6 (hc. : alk. paper)
1. Science—Statistical methods. 2. Monte Carlo method. I. Title. II. Series
Q180.55.S7 L58 2001
501'519282—dc21 00-069243

ISBN 0-387-95230-6

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Contents

1

Introduction and Examples

1.1 The Need of Monte Carlo Techniques

An essential part of many scientific problems is the computation of integral

$$I = \int_D g(\mathbf{x}) d\mathbf{x},$$

where D is often a region in a high-dimensional space and $g(\mathbf{x})$ is the target function of interest. If we can draw independent and identically distributed (i.i.d.) random samples $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ uniformly from D (by a computer), an approximation to I can be obtained as

$$\hat{I}_m = \frac{1}{m}\{g(\mathbf{x}^{(1)}) + \cdots + g(\mathbf{x}^{(m)})\}.$$

The *law of large numbers* states that the average of many independent random variables with common mean and finite variances tends to stabilize at their common mean (see the Appendix); that is,

$$\lim_{m\to\infty} \hat{I}_m = I, \quad \text{with probability 1.}$$

Its convergence rate can be assessed by the *central limit theorem* (CLT):

$$\sqrt{m}(\hat{I}_m - I) \to N(0, \sigma^2), \quad \text{in distribution,}$$

where $\sigma^2 = \text{var}\{g(\mathbf{x})\}$. Hence, the "error term" of this Monte Carlo approximation is $O(m^{-1/2})$, regardless of the dimensionality of $\mathbf{x}$. This basic setting underlies the potential role of the Monte Carlo methodology in science and statistics.

In the simplest case when $D = [0,1]$ and $I = \int_0^1 g(x)dx$, one can approximate I by

$$\tilde{I}_m = \frac{1}{m}\left\{g(b_1) + \cdots + g(b_m)\right\},$$

where $b_j = j/m$. This method can be called the *Riemann approximation.* When g is reasonably smooth, the Riemann approximation gives us an error rate of $O(m^{-1})$, better than that of the Monte Carlo method. More sophisticated methods such as Simpson's rule and the Newton-Cotes rules give better numerical approximations (Thisted 1988). However, a fatal defect of these deterministic methods is that they do not scale well as the dimensionality of D increases. For example, in a 10-dimensional space with $D = [0,1]^{10}$, we will have to evaluate $O(m^{10})$ grid points in order to achieve an accuracy of $O(m^{-1})$ in the Riemann approximation of I. In contrast, the naive Monte Carlo approach, which draws $x^{(1)}, \ldots, x^{(m)}$ uniformly from D, has an error rate $O(m^{-1/2})$ regardless of the dimensionality of D, at least theoretically.

Although the "error rate" of a Monte Carlo integration scheme remains the same in high-dimensional problems, two intrinsic difficulties arise: (a) when the region D is large in high-dimensional space, the variance σ^2, which measures how "uniform" the function g is in region D, can be formidably large; (b) one may not be able to produce uniform random samples in an arbitrary region D. To overcome these difficulties, researchers often employ the idea of *importance sampling* in which one generates random samples $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ from a nonuniform distribution $\pi(\mathbf{x})$ that puts more probability mass on "important" parts of the state space D. One can estimate integral I as

$$\hat{I} = \frac{1}{m}\sum_{j=1}^{m}\frac{g(\mathbf{x}^{(j)})}{\pi(\mathbf{x}^{(j)})},$$

which has a variance $\sigma_\pi^2 = \text{var}_\pi\{g(\mathbf{x})/\pi(\mathbf{x})\}$. In the most fortunate case, we may choose $\pi(\mathbf{x}) \propto g(\mathbf{x})$ when g is non-negative and I is finite, which results in an exact estimate of I. But in no known application of the Monte Carlo method has this "luckiest situation" ever occurred. More realistically, we may hope to find a good "candidate" π which will explore more in regions where the value of g is high. In such a situation, generating random draws from π can be a challenging problem.

Demands for sampling from a nonuniform distribution π are also seen from another set of problems in bioinformatics, computational chemistry, physics, structural biology, statistics, etc. In these problems, the desired probability distribution $\pi(\mathbf{x})$ of a complex system, where $\mathbf{x}$ is often called a *configuration* of the system, arises from basic laws in physics and statistical inference. For example, in the study of a macromolecule, $\mathbf{x}$ may represent the *structure* of a molecule in the form of three-dimensional coordinates of all the atoms in the molecule. The target probability distribution is defined by the Boltzmann distribution $\pi(\mathbf{x}) = Z(T)e^{-h(\mathbf{x})/kT}$, where k is the Boltzmann constant, T is the system's temperature, $h(\mathbf{x})$ is the energy function, and $Z(T)$ is the *partition function* which is difficult to compute. Scientists are often interested in certain "average characteristics" of the system, many of which can be expressed mathematically as $E_\pi[g(\mathbf{x})]$ for a suitable function g. In Bayesian statistical inference, $\mathbf{x}$ often represents the joint configuration of missing data and parameter values and $\pi(\mathbf{x})$ is usually the posterior distribution of these variables. One has to integrate out nuisance parameters and the missing data so as to make a proper inference on the parameter of interest and to make valid predictions for future observations. These tasks can, once again, be expressed as computing the expectation of a function of the configuration space.

Sometimes, an optimization problem can also be formulated as a Monte Carlo sampling problem. Suppose we are interested in finding the minimum of a target function, $h(\mathbf{x})$, defined on a possibly complex configuration space. The problem is equivalent to finding the maximum of another function, $q_T(\mathbf{x}) = e^{-h(\mathbf{x})/T}$ (as long as $T > 0$). In the case when $q_T(\mathbf{x})$ is integrable for all $T > 0$, which is most common in practice, we can make up a family of probability distributions:

$$\pi_T(\mathbf{x}) \propto e^{-h(\mathbf{x})/T}, \quad T > 0.$$

If we can sample from $\pi_T(\mathbf{x})$ when T is sufficiently small, resulting random draws will most likely be located in the vicinity of the global minimum of $h(\mathbf{x})$. This consideration is the basis of the well-known simulated annealing algorithm (Kirkpatrick et al. 1983) and is also key to the *tempering* techniques for designing more efficient Monte Carlo algorithms (Chapter 10).

1.2 Scope and Outline of the Book

A fundamental step in all Monte Carlo methods is to generate (pseudo-) random samples from a probability distribution function $\pi(\mathbf{x})$, often known only up to a normalizing constant. The variable of interest $\mathbf{x}$ usually takes value in $\mathbb{R}^k$, but occasionally can take value in other spaces such as a permutation or transformation group (Diaconis 1988, Liu and Wu 1999). In most applications, directly generating independent samples from the target distribution π is not feasible. It is often the case that either the generated samples have to be dependent or the distribution used to generate the samples is different from π, or both. The rejection method (von Neumann 1951), importance sampling (Marshall 1956), and sampling-importance-resampling (SIR) (Rubin 1987) are schemes that make use of samples generated from a *trial distribution* $p(\mathbf{x})$, which differs from, but should be similar to, the target distribution π. The Metropolis algorithm (Metropolis et al. 1953) which, together with Hastings's (1970) generalizations, serves as the basic building block of Markov chain Monte Carlo (MCMC), is the one that generates dependent samples from a Markov chain with π as its equilibrium distribution. In other words, MCMC is essentially a Monte Carlo integration procedure in which the random samples are produced by evolving a Markov chain.

Because of the great potential of Monte Carlo methodology, various techniques have been developed by researchers in their respective fields. Recent advances in Monte Carlo techniques include the cluster method, data augmentation, parameter expansion, multicanonical sampling, multigrid Monte Carlo (MGMC), umbrella sampling, density-scaling Monte Carlo, simulated tempering, parallel tempering, hybrid Mont Carlo (HMC), multiple try Metropolis (MTM), sequential Monte Carlo, particle filtering, etc. There is also a trend in moving toward a population-based approach. These advances in one way or another were all motivated by the need to sample from very complex probability distributions for which the standard Metropolis method tends to be trapped in a local "energy" well. Many of these methods are related, and some are even identical. For example, the configurational bias Monte Carlo (Siepmann and Frenkel 1992) is equivalent to a sequential importance sampling combined with a Metropolized independence sampler (Chapters 2 & 3); the exchange Monte Carlo (Hukushima and Nemoto 1996) is reminiscent of *parallel tempering* (Geyer 1991); the multiple-try Metropolis (Liu, Liang and Wong 2000) generalizes a method described by Frenkel and Smit (1996); the parameter expansion (Liu and Wu 1999) recently developed is a special case of the *partial resampling* technique (Goodman and Sokal 1989); and the bootstrap filter and sequential imputation (Gordon, Salmond and Smith 1993, Kong, Liu and Wong 1994) can be traced back to the "growth method" (Hammersley and Morton 1954, Rosenbluth and Rosenbluth 1955). By providing a systematic account of these methods, this book focuses on the following aspects: understanding the properties and characteristics of these methods, revealing their connections and differences, comparing their performances and proposing generalizations, and demonstrating their use in scientific and statistical problems.

The remaining part of this chapter presents motivating examples in statistical physics, molecular simulation, bioinformatics, dynamic system analysis, statistical hypothesis testing, Bayesian inference for hierarchical models, and other statistical missing data problems.

Chapter 2 covers basic Monte Carlo techniques including the inversion method, rejection sampling, antithetic sampling, control variate method, stratified sampling, importance sampling, and the exact sampling method for chain-structured models. The last method is usually not covered by the standard Monte Carlo books but is becoming increasingly important in