

PREFACE

The technique of Markov chain Monte Carlo (MCMC) first arose in statistical physics, marked by the celebrated 1953 paper of Metropolis, Rosenbluth, Rosenbluth, Teller and Teller. The underlying principle is simple: if one wishes to sample randomly from a specific probability distribution then design a Markov chain whose long-time equilibrium is that distribution, write a computer program to simulate the Markov chain, and run the programmed chain for a time long enough to be confident that approximate equilibrium has been attained; finally record the state of the Markov chain as an approximate draw from equilibrium. The Metropolis *et al.* paper used a symmetric Markov chain; later developments included adaptation to the case of non-symmetric Markov chains.

The technique has developed strongly in the statistical physics community but also in separate ways and with rather different emphases in the computer science community concerned with the study of random algorithms (where the emphasis is on whether the resulting algorithm scales well with increasing size of the problem), in the spatial statistics community (where one is interested in understanding what kinds of patterns arise from complex stochastic models), and also in the applied statistics community (where it is applied largely in Bayesian contexts, enabling researchers to formulate statistical models which would otherwise be intransigent to effective statistical analyses).

Within the statistical physics community, the MCMC technique lies at the heart of the tradition of “simulation physics”: understanding phase transition and other physical behaviour by constructing careful simulation experiments on the computer. A particular line of development for the past 10 years in the statistical physics community is that of extended ensemble methods, beginning with Berg’s work on the multicanonical method, followed by simulated tempering, parallel tempering, broad histogram Monte Carlo, transition matrix Monte Carlo, etc. These methods substantially

extend the ability to simulate complicated systems that are very difficult to deal with directly, such as spin-glasses, or protein models.

Within the statistics community, landmark papers include the famous Geman–Geman 1984 paper on image restoration, work by Gelfand and Smith in 1990 showing that MCMC can be applied effectively to Bayesian problems, and Green’s (1995) work on dimension-varying problems. The resulting impact on applied statistics has been truly revolutionary.

A recent theoretical development is that of *perfect simulation*, addressing the following central question: how long should one run the Markov chain so as to ensure that it is close to equilibrium? This rather startling development is as follows: in favourable cases one can adjust the Markov chain Monte Carlo algorithm so as to generate exact draws from the target distribution. It was given practical effect in two different ways in the seminal papers of Propp and Wilson (1996) and of Fill (1998). Subsequent work has filled out the mathematical picture by clarifying how recent developments relate to previous work (both the Propp–Wilson and Fill algorithms relate in interesting ways to each other and to prior theoretical concepts, while adding their own attractively empirical flavour).

The development of theory also benefits applications: simulation techniques have been applied to develop practical statistical inferences for almost all problems in (bio)statistics, for example, the problems in longitudinal data analysis, image analysis, genetics, contagious disease epidemics, random spatial pattern, and financial statistical models such as GARCH and stochastic volatility. The techniques also constitute a major part of today’s bioinformatics toolbox.

The expositions which make up this book arose from a desire to bring together people who work on innovative developments and applications across the range of statistics, physics, and bioinformatics, to encourage cross-fertilization and to challenge each other with varied problems. The Institute of Mathematical Sciences of the National University of Singapore kindly and generously agreed to fund a month-long programme of activity in March 2004, which allowed us to invite a number of distinguished lecturers from the different fields, to present courses at graduate level; this resulted in a memorable and most productive month, greatly facilitated by the kindness and efficiency of the IMS Director, Prof Louis Chen, and his talented and able staff. The chapters of this book correspond to several of these courses: in Chapter 1 Bernd Berg introduces Markov chain Monte Carlo from the perspective of statistical physics, starting from simple ideas of probability, and developing MCMC ideas right up to multicanonical en-

sembles, illustrated using FORTRAN computer code available on the web. Chapter 2 presents a complementary view from David Landau, with particular emphasis on issues of finite-size effects, the peculiarities of random number generators, and a spectrum of ingenious techniques to assess phase transition effects. In a change of pace, Wilfrid Kendall uses Chapter 3 to describe the various ideas involved in the transformations of algorithms known collectively as perfect simulation, which in favourable cases deliver exact draws in random rather than deterministic runtimes. A very different theme is treated by Rong Chen in Chapter 4: simulation algorithms that process information sequentially (known as Sequential Monte Carlo), either because this is natural to the algorithm itself, or because it is useful to decompose the problem in such a manner. Finally, in Chapter 5, Elizabeth Thompson presents a careful study of how MCMC is put into practice in the analysis of pedigrees in genetics.

It is our hope as editors that this ensemble of expositions, and the diversity of ideas contained therein, will form an attractive invitation to readers, to introduce them to the fascinating and various worlds of Markov chain Monte Carlo in mathematical science. We trust you will enjoy this book as much as we enjoyed the task of its compilation!

W. S. Kendall
F. Liang
J.-S. Wang