Monte Carlo Methods - Fall 2013/14

Goethe Universität, Frankfurt am Main

Lecture/Tutorial: Wolfgang Unger Office: 02.105 unger@th.physik.uni-frankfurt.de

---

Exercise Nr. 5

Discussion on November 18th, 14:00-15:00

13) Metropolis-Hastings algorithm *(5+3 points)*

Generate a binomial distribution $\pi(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$ via the Metropolis-Hastings algorithm as described in the lecture. Plot $\pi(k)$, and compute the mean value and the variance for the probabilities (a) $p = 1/2$, (b) $p = 1/3$, (c) $p = 1/4$, and in each case for $n = 4, 8, 16, 32$.

14) Ising Chain *(3+3 points)*

Study the Ising chain for the ferromagnetic ($J = 1$) and anti-ferromagnetic ($J = -1$) model.

(a) Implement either the Metropolis algorithm or the Heatbath algorithm and measure the energy density.

(b) Add to the Hamiltonian an external field h coupling to the spins:

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i.$$

Derive the analytic solution in this more general case, by diagonalizing the corresponding transfer matrix and determine the energy density. As an intermediate step, verify that the eigenvalues are

$$\lambda_\pm = e^{\beta J} \left(\cosh(\beta h) \pm \sqrt{\sinh^2(\beta h) + e^{-4\beta J}} \right)$$

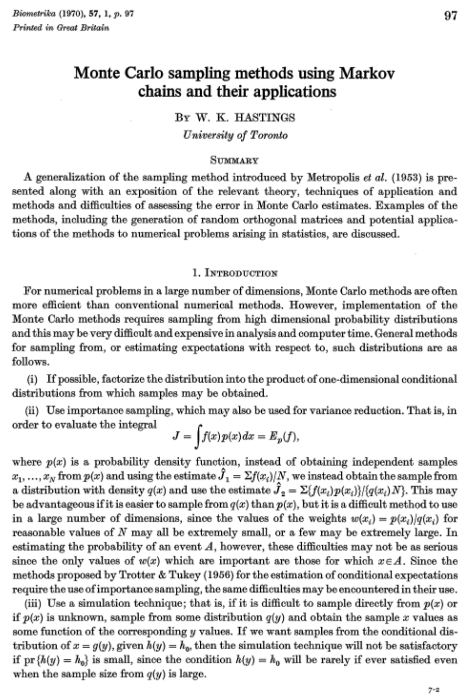
(*c) *Extra points (+3):* incorporate the external magnetic field in your simulation and measure the energy. Does it agree with the analytic result in the limit of large N?

15) Random Walk *(2+2+2)*

Implement a random walk (RW) on (a) a 1-dimensional chain, (b) a square lattice and (c) a cubic lattice and measure the root mean square distance between the starting point at the origin and the end point after N steps. For a statistical analysis you have to sample many walkers. Does your critical exponent agree with $\nu = 2$ within statistical errors?

W. Keith Hastings

(July 21, 1930 - *) is an Canadian statistician.


Biometrika (1970), 57, 1, p. 97
Printed in Great Britain

97

Monte Carlo sampling methods using Markov chains and their applications

By W. K. HASTINGS

University of Toronto

SUMMARY

A generalization of the sampling method introduced by Metropolis *et al.* (1953) is presented along with an exposition of the relevant theory, techniques of application and methods and difficulties of assessing the error in Monte Carlo estimates. Examples of the methods, including the generation of random orthogonal matrices and potential applications of the methods to numerical problems arising in statistics, are discussed.

1. INTRODUCTION

For numerical problems in a large number of dimensions, Monte Carlo methods are often more efficient than conventional numerical methods. However, implementation of the Monte Carlo methods requires sampling from high dimensional probability distributions and this may be very difficult and expensive in analysis and computer time. General methods for sampling from, or estimating expectations with respect to, such distributions are as follows.

(i) If possible, factorize the distribution into the product of one-dimensional conditional distributions from which samples may be obtained.

(ii) Use importance sampling, which may also be used for variance reduction. That is, in order to evaluate the integral

$$J = \int f(x)p(x)dx = E_p(f),$$

where $p(x)$ is a probability density function, instead of obtaining independent samples $x_1, \ldots, x_N$ from $p(x)$ and using the estimate $\hat{J}_1 = \Sigma f(x_i)/N$, we instead obtain the sample from a distribution with density $q(x)$ and use the estimate $\hat{J}_2 = \Sigma\{f(x_i)p(x_i)\}/\{q(x_i)N\}$. This may be advantageous if it is easier to sample from $q(x)$ than $p(x)$, but it is a difficult method to use in a large number of dimensions, since the values of the weights $w(x_i) = p(x_i)/q(x_i)$ for reasonable values of N may all be extremely small, or a few may be extremely large. In estimating the probability of an event A, however, these difficulties may not be as serious since the only values of $w(x)$ which are important are those for which $x \in A$. Since the methods proposed by Trotter & Tukey (1956) for the estimation of conditional expectations require the use of importance sampling, the same difficulties may be encountered in their use.

(iii) Use a simulation technique; that is, if it is difficult to sample directly from $p(x)$ or if $p(x)$ is unknown, sample from some distribution $q(y)$ and obtain the sample x values as some function of the corresponding y values. If we want samples from the conditional distribution of $x = g(y)$, given $h(y) = h_0$, then the simulation technique will not be satisfactory if pr$\{h(y) = h_0\}$ is small, since the condition $h(y) = h_0$ will be rarely if ever satisfied even when the sample size from $q(y)$ is large.

7-2

The Metropolis-Hastings algorithm (or, Hastings-Metropolis algorithm) is the most common Markov chain Monte Carlo (MCMC) method. It is extremely widely used in applied statistics (and in statistical physics and computer science), to sample from complicated, high-dimensional probability distributions. A primary source for this algorithm is the paper [right]. This paper has been cited well over two thousand times - a huge number. However, despite this paper's importance, very little information about W.K. Hastings himself is publicly available. [...] W. Keith Hastings was born on July 21, 1930, in Toronto, Ontario, Canada. He received his B.A. in Applied Mathematics from the University of Toronto in 1953, and then worked from 1955-59 as a "Consultant in Computer Applications" for the Toronto company H.S. Gellman & Co. Hastings recalls:

> Harvey Gellman was a good mentor and encouraged me to pursue my ideas. Some of the projects involved simulations and this was my first contact with statistics and generation of samples from probability distributions.

Overlapping somewhat with this, Hastings received his M.A. in 1958, and his Ph.D. in 1962, both from the University of Toronto's Department of Mathematics (which included Statistics at that time). His Ph.D. thesis title was "Invariant Fiducial Distributions". His Ph.D. supervisor was initially Don Fraser (who mentioned Hastings' thesis results in a January 10, 1962 letter to R.A. Fisher), and later Geoffrey Watson (while Fraser visited Stanford in 1961-62). After completing his Ph.D., Hastings worked briefly at the University of Canterbury in New Zealand (1962-64), and at Bell Labs in New Jersey (1964-66). Hastings writes:

> I was never comfortable working on statistical inference for my thesis. My investigations led to too many dead ends and the work seemed to involve more mathematical considerations than statistical ones. When Geoff took over as my supervisor I briefly considered changing topics, but ended up sticking with my original topic and completed my thesis. In New Zealand, I continued this work for a while but eventually gave it up, the final blow coming when I learned that Fiducial Probability was declared 'dead' in a session during a statistics conference held in Ottawa. Bell Labs provided a welcome and effective antidote to all this as I gradually turned towards the computational aspects of statistics. In effect, I was then returning to my professional roots.

From 1966 to 1971, Hastings was an Associate Professor in the Department of Mathematics at the University of Toronto. During this period, he wrote the famous paper listed above (which generalised the work of N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), "Equations of state calculations by fast computing machines", J. Chem. Phys. 21, 1087-1091). Hastings explains:

> When I returned to the University of Toronto, after my time at Bell Labs, I focused on Monte Carlo methods and at first on methods of sampling from probability distributions with no particular area of application in mind. [University of Toronto Chemistry professor] John Valleau and his associates consulted me concerning their work. They were using Metropolis's method to estimate the mean energy of a system of particles in a defined potential field. With 6 coordinates per particle, a system of just 100 particles involved a dimension of 600. When I learned how easy it was to generate samples from high dimensional distributions using Markov chains, I realised how important this was for Statistics, and I devoted all my time to this method and its variants which resulted in the 1970 paper.

[...] In 1971, Hastings joined the Department of Mathematics at the University of Victoria. [...] Hastings retired from the University of Victoria in 1992. As of this writing, he still lives in Victoria.

`[from Jeffrey S. Rosenthal, http://probability.ca/hastings]`