Monte Carlo Methods

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu
Web site: www.physics.cornell.edu/∼cyrus

Feb 15,17, 2010 to Jim Sethna's numerical methods class

Outline

1. Why use Monte Carlo? Early history of MC.
2. Monte Carlo vs. deterministic integration
3. Importance sampling
4. Pseudo-random vs. quasi-random numbers
5. Sampling nonuniform probability densities
    1. Transformation method
    2. Rejection method
    3. Metropolis-Hastings algorithm

When to use Monte Carlo Methods

Monte Carlo methods: A class of computational algorithms that rely on repeated random sampling to compute results.

A few broad areas of applications are:

1. physics
2. chemistry
3. engineering
4. finance and risk analysis

When are MC methods likely to be the methods of choice?

1. When the problem is many-dimensional and approximations that factor the problem into products of lower dimensional problems are inaccurate.
2. A less important reason is that if one has a complicated geometry, a MC algorithm may be simpler than other choices.

Obvious drawback of MC methods: There is a statistical error. Sometimes there is a tradeoff between statistical error and systematic error and one needs to find the best compromise.

Quantum Monte Carlo

Quantum Monte Carlo: Monte Carlo methods used to solve the Schrödinger Eq.
Some systems to which they have been applied are:

- strongly correlated systems (Hubbard, Anderson, t-J, ... models)
- quantum spin systems (Ising, Heisenberg, xy, ... models),
- liquid and solid helium, liquid-solid interface, droplets
- energy and response of homogeneous electron gas in 2-D and 3-D
- nuclear structure
- lattice gauge theory
- atomic clusters
- electronic structure calculations of atoms, molecules and solids

- both to pure states and finite temperature problems

MC Simulations versus MC calculations

One can distinguish between two kinds of algorithms:

1. The system being studied is stochastic and the stochasticity of the algorithm mimics the stochasticity of the actual system. e.g. study of neutron transport and decay in nuclear reactor by following the trajectories of a large number of neutrons. Such problems are suitable for MC algorithms in a very obvious way.
2. Much more interesting are applications where the system being studied is not stochastic, but nevertheless a stochastic algorithm is the most efficient, or the most accurate, or the only feasible method for studying the system. e.g. the solution of a PDE in a large number of variables, e.g., the solution of the Schrödinger equation for an N-electron system, with say $N = 100$ or 1000. (Note: The fact that the wavefunction has a probabilistic interpretation has *nothing* to do with the stochasticity of the algorithm. The wavefunction itself is perfectly deterministic.)

I prefer to use the terminology that the former are MC simulations whereas the latter are MC calculations but not everyone abides by that terminology.

Early Recorded History of Monte Carlo

1777 Compte de Buffon: If a needle of length L is thrown at random onto a plane ruled with straight lines a distance $d(d > L)$ apart, then the probability P of the needle intersecting one of those lines is $P = \frac{2L}{\pi d}$.
Laplace: This could be used to compute π (inefficiently).

1930s First significant scientific application of MC: Enrico Fermi used it for neutron transport in fissile material.
Segre: "Fermi took great delight in astonishing his Roman colleagues with his "too-good-to-believe" predictions of experimental results."

1940s Monte Carlo named by Nicholas Metropolis and Stanislaw Ulam

1953 Algorithm for sampling any probability density
Metropolis, Rosenbluth, Rosenbluth, Teller and Teller
(generalized by Hastings in 1970)

1962,1974 First QMC calculations, Kalos, and, Kalos, Levesque, Verlet.
1965 First VMC calculations (of liquid He), Bill McMillan.

Monte Carlo versus Deterministic Integration methods

Deterministic Integration Methods:
Integration Error, ϵ, using $N_{\rm int}$ integration points:
1-dim Simpson rule: $\epsilon \propto N_{\rm int}^{-4}$, (provided derivatives upto 4^{th} exist)
d-dim Simpson rule: $\epsilon \propto N_{\rm int}^{-4/d}$, (provided derivatives upto 4^{th} exist)
So, for a given error, N and so the computer time increases exponentially with d, since $N \propto (\frac{1}{\epsilon})^{d/4}$.

Monte Carlo:
$\epsilon \propto N_{\rm int}^{-1/2}$, independent of dimension!, according to the central limit theorem provided that the variance of the integrand is finite.

Roughly, Monte Carlo becomes advantageous for $d > 8$.
For a many-body wavefunction $d = 3N$ and can be a few thousand!
Remarkably, by the law of large numbers, even when the variance is infinite, if the expected value is finite, the MC estimate will converge, but more slowly than $N^{-1/2}$.

Monte Carlo Integration

$$I \; = \; \int_V f(x)dx \;\; = \;\; V\langle f\rangle \pm V\sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N-1}}$$

$$\text{where} \quad \langle f\rangle \; = \; \frac{1}{N}\sum_i^N f(x_i), \quad \langle f^2\rangle = \frac{1}{N}\sum_i^N f^2(x_i)$$

and the points x_i are sampled uniformly in V.

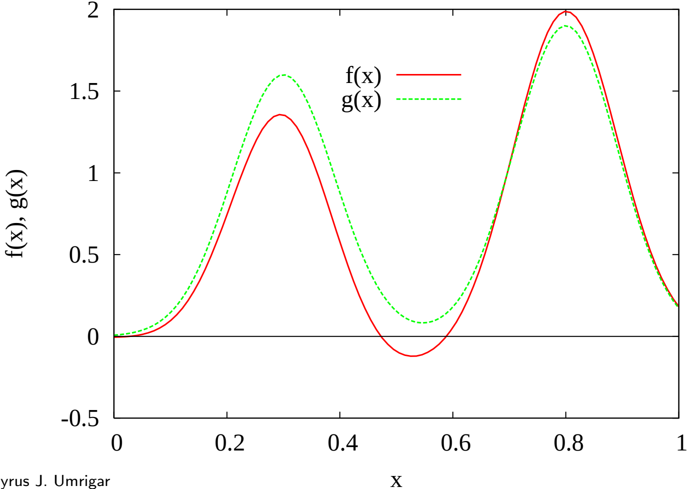
Importance sampling

$$I \; = \; \int_V g(x)\frac{f(x)}{g(x)}dx \;\; = \;\; V\left\langle\frac{f}{g}\right\rangle \pm V\sqrt{\frac{\left\langle\left(\frac{f}{g}\right)^2\right\rangle - \left\langle\frac{f}{g}\right\rangle^2}{N-1}}$$

where the probability density function $g(x) \geq 0$ and $\int_V g(x)dx = 1$.
If $g(x) = 1/V$ in V then we recover original fluctuations but if $g(x)$ mimics $f(x)$ then the fluctuations are much reduced. Need: a) $g(x) \geq 0$, b) know integral of $g(x)$, and, c) be able to sample it.

Illustration of Importance Sampling

$f(x)$ is the function to be integrated. $g(x)$ is a function that is "similar" to $f(x)$ and has the required properties: a) $g(x) \geq 0$, b) we know integral of $g(x)$, and, c) we know how to sample it. $\int f(x)dx$ can be evaluated efficiently by sampling $g(x)$ and averaging $f(x)/g(x)$.



Example of Importance Sampling to Calculate Integrals More Efficiently

Suppose we wish to compute

$$\int_0^1 dx\, f(x) \;=\; \int_0^1 dx\, \frac{1}{\sqrt{x}+x} \quad \text{(Exactly integrable, but pretend not)}$$

Note that

$$\int_0^1 dx(f(x))^2 \;=\; \infty, \quad \text{(Barely, log divergence)}$$

so if we estimate the integral by sampling points uniformly in $[0,1]$ then this would be an infinite variance estimator and the error of the estimate will go down more slowly than $N^{-1/2}$. However, we can instead sample points from the density

$$g(x) \;=\; \frac{1}{2x^{1/2}}$$

Now the variance of $f(x)/g(x)$ is finite and the error decreases as $N^{-1/2}$, and, with a small prefactor. (Still would not use this in 1D.)

Estimating Unbiased Variance from Samples

Let $\langle f(x) \rangle$ denote the population mean and $\bar{f}(x)$ denote the sample mean. Then

$$\left\langle \frac{\sum_i f^2(x_i)}{N} - \left[\frac{\sum_i f(x_i)}{N} \right]^2 \right\rangle = \langle f^2 \rangle - \left\langle \frac{\sum_i f^2(x_i) + \sum_{i,i \neq j} \sum_j f(x_i) f(x_j)}{N^2} \right\rangle$$

Since $f(x_i)$ and $f(x_j)$ are independent

$$RHS = \left(1 - \frac{1}{N}\right) \langle f^2 \rangle - \frac{N(N-1)}{N^2} \langle f \rangle^2 \; = \; \frac{N-1}{N} (\langle f^2 \rangle - \langle f \rangle^2) \; = \; \frac{N-1}{N} \sigma^2$$

So, the sample estimate for σ^2 is

$$\boxed{\sigma^2 \approx \frac{N}{N-1} \left(\bar{f^2} - (\bar{f})^2 \right)}$$

Loss of one degree of freedom because sample variance is computed relative to sample mean rather than the true mean.

Infinite variance estimators

When variance σ^2 is finite, by the central limit theorem the average

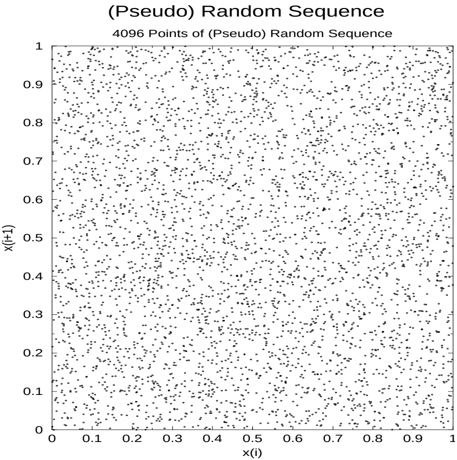
$$F_N = \frac{\sum_{i=1}^{N} f(x_i)}{N}$$

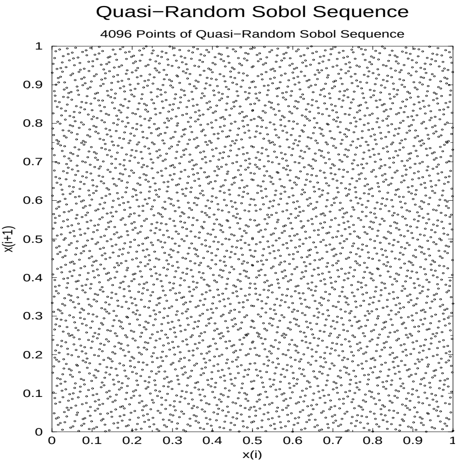
converges for increasing N to a gaussian of width $\sigma_N = \sigma/\sqrt{N}$.
Since we have a gaussian distribution the probability of
F_n being within $1\sigma_N$ of the true mean is 68.3%
F_n being within $2\sigma_N$ of the true mean is 95.4%
F_n being within $3\sigma_N$ of the true mean is 99.7%.

What if the population variance $\sigma^2 = \infty$ but we do not know that beforehand? The computed sample variance will ofcourse always be finite. So, how do we know if we expect the error of the sample mean to go down as $\frac{1}{\sqrt{N}}$ or more slowly?
The practical signature of an infinite variance estimator is that when one computes σ from increasingly larger samples, the estimate of the variance will have large upward jumps.

Pseudo-random vs quasi-random numbers

Terrible misnomers!





Reason why uniform grid is inefficient: Projection of $N = n^d$ points in d dimensions onto a line maps n^{d-1} points onto a single point.

Reason why quasi-MC is more efficient than pseudo-MC in intermediate dimensions: Quasi-MC avoids clusters and voids.

Random Number Generators

Conventional random number generators generate random numbers uniformly distributed on [0,1).

Of course no computer generated sequence of random numbers is truly random. For one, the random numbers must repeat after a finite (though hopefully very large) period. Also, if N bits are used to represent the random numbers, then the number of different numbers generated can by no larger than 2^N.

Note however, that the period can be (and typically is for the better generators) much larger than 2^N.

Many different algorithms exist for generating random numbers, e.g., linear congruential generators (with or without an additive constant), linear feedback shift register, lagged Fibonacci generator, XORshift algorithm etc. They are typically subjected to a battery of statistical tests, e.g., the Diehard tests of Marsaglia. Of course no random number generator can pass all the tests that one can invent, but hopefully the random number generator used does not have correlations that could significantly impact the system being studied.

Random Number Generators

For many MC calculations it is the short-ranged correlations that matter most, but one has to think for each application what is important. For example, if one were studying an Ising model with a power of two number of spins, it would be problematic to have random number generator that generated numbers with bits that repeat at an interval of 2^N.

In the old days, there were quite a few calculations that produced inaccurate results due to bad random number generators. For example, the standard generators that came with UNIX and with C were badly flawed. In the 1980s a special purpose computer was built at Santa Barbara to study the 3-D Ising model. However, at first it failed to reproduce the known exact results for the 2-D Ising model and that failure was traced back to a faulty random number generator.

Sampling random variables from nonuniform probability density functions

We say x is sampled from $f(x)$ if for any a and b in the domain,

$$\mathrm{Prob}[a \leq x \leq b] \; = \; \int_a^b dx' \; f(x')$$

1) Transformation method (For many simple functions)
2) Rejection method (For more complicated functions)
3) Metropolis-Hastings method (For any function)

1) Transformation method: Perform a transformation $y(x)$ on a uniform deviate x, to get y sampled from desired probability density $f(y)$.

$$|P(x)dx| \; = \; |P(y)dy| \quad \text{conservation of probability}$$

If we have sampled x from a uniform density ($P(x) = 1$) and we wish y to be sampled from the desired density, $f(y)$, then setting $P(y) = f(y)$,

$$\left|\frac{dx}{dy}\right| \; = \; f(y)$$

Solve for $x(y)$ and invert to get $y(x)$, i.e., invert the cumulative distrib.

Examples of Transformation Method

Example 1: $f(y) = ae^{-ay}$, $y \in [0, \infty)$

$$\left|\frac{dx}{dy}\right| \quad = \quad ae^{-ay}, \quad \text{or,} \quad x = e^{-ay}, \quad \text{i.e.,} \quad \boxed{y = \frac{-\ln(x)}{a}}$$

Example 2: $f(y) = \frac{y^{-1/2}}{2}$, $y \in [0, 1]$

$$\left|\frac{dx}{dy}\right| \quad = \quad \frac{y^{-1/2}}{2}, \quad \text{or} \quad x = y^{1/2}, \quad \text{i.e.,} \quad \boxed{y = x^2}$$

Note that in this case we are sampling a probability density that is infinite at 0, but that is OK!

Examples of Transformation Method

Example 3: $f(y) = ye^{-y^2/2}$, $y \in [0, \infty)$

$$\left|\frac{dx}{dy}\right| = ye^{-y^2/2}, \quad \text{or,} \quad x = e^{-y^2/2}, \quad \text{i.e.,} \quad \boxed{y = \sqrt{-2\ln(x)}}$$

Example 4a: $f(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}$, $y \in (-\infty, \infty)$ (using Box-Müller method)

$$\frac{1}{2\pi} e^{-(\frac{y_1^2}{2}+\frac{y_2^2}{2})}\, dx\, dy = \left(r\, e^{-\frac{r^2}{2}}\, dr\right)\left(\frac{d\phi}{2\pi}\right)$$

$$r = \sqrt{-2\log(x_1)}, \qquad \phi = 2\pi x_2$$

$$\boxed{y_1 = \sqrt{-2\log(x_1)}\cos(2\pi x_2)}, \qquad \boxed{y_2 = \sqrt{-2\log(x_1)}\sin(2\pi x_2)}$$

(y_1 and y_2 are uncorrelated)

Example 4b: $f(y) \approx \frac{e^{-y^2/2}}{\sqrt{2\pi}}$, $y \in (-\infty, \infty)$ (using central-limit theorem)

$$y = \lim_{N\to\infty} \sqrt{\frac{12}{N}} \left(\sum_{i=1}^{N} x_i - \frac{N}{2}\right) \approx \boxed{\sum_{i=1}^{12} x_i - 6}$$

(avoids sqrt, log, cos, sin, but, misses tiny tails beyond ±6)

Rejection Method

We wish to sample $f(x)$.
Find a function $g(x)$ that can be sampled by another method (say transformation) and that preferably mimics the behaviour of $f(x)$.
Let C be an upper bound to the maximum value of $f(x)/g(x)$.
Let $C \geq \max(f(x)/g(x))$.
Then $f(x)$ is sampled by sampling $g(x)$ and keep the sampled points with probability

$$P = \frac{f(x)}{Cg(x)}$$

The efficiency of the method is the fraction of the sampled points that are kept.

$$\begin{aligned} Eff &= \int dx \frac{f(x)}{Cg(x)} g(x) \\ &= \frac{1}{C} \end{aligned}$$

Drawback: It is often hard to know C and a “safe” upperbound choice for C may lead to low efficiency. An alternative is to associate weights with the sampled points.

Variational Monte Carlo

W. L. McMillan, Phys. Rev. **138**, A442 (1965)

Monte Carlo is used to perform the many-dimensional integrals needed to calculate quantum mechanical expectation values. e.g.

$$
\begin{aligned}
E &= \frac{\int d\mathbf{R}\, \Psi_{\mathrm{T}}^*(\mathbf{R})\, \mathcal{H}\, \Psi_{\mathrm{T}}(\mathbf{R})}{\int d\mathbf{R}\, |\psi_{\mathrm{T}}(\mathbf{R})|^2} \\
&= \int d\mathbf{R}\; \frac{|\psi_{\mathrm{T}}(\mathbf{R})|^2}{\int d\mathbf{R}\, |\psi_{\mathrm{T}}(\mathbf{R})|^2}\; \frac{\mathcal{H}\Psi_{\mathrm{T}}(\mathbf{R})}{\Psi_{\mathrm{T}}(\mathbf{R})} \\
&= \frac{1}{N}\sum_i \frac{\mathcal{H}\Psi_{\mathrm{T}}(\mathbf{R}_i)}{\Psi_{\mathrm{T}}(\mathbf{R}_i)} = \frac{1}{N}\sum_i E_L(\mathbf{R}_i)
\end{aligned}
$$

Energy is obtained as an arithmetic sum of the *local energies* $E_L(\mathbf{R}_i)$ evaluated for configurations sampled from $|\psi_{\mathrm{T}}(\mathbf{R})|^2$ using a generalization of the Metropolis method. If ψ_{T} is an eigenfunction the $E_L(\mathbf{R}_i)$ do not fluctuate. Accuracy of VMC depends crucially on the quality of $\Psi_{\mathrm{T}}(\mathbf{R})$. Diffusion MC does better by projecting onto ground state.

Three ingredients for accurate Variational Monte Carlo

1. A method for sampling an arbitrary wave function Metropolis-Hastings.
2. A functional form for the wave function that is capable of describing the correct physics/chemistry. Beyond the scope of these lectures.
3. An efficient method for optimizing the parameters in the wave functions. Beyond the scope of these lectures.

Metropolis-Hastings Monte Carlo

Metropolis, Rosenbluth^2, Teller^2, JCP, **21** 1087 (1953)
W.K. Hastings, Biometrika, **57** (1970)

Metropolis method originally used to sample the Boltzmann distribution. This is still one of its more common uses.

General method for sampling **any known** discrete or continuous density. (Other quantum Monte Carlo methods, e.g., diffusion MC, enable one to sample densities that are not explicitly known but are the eigenstates of known matrices or integral kernels.)

Metropolis-Hastings has serial correlations. Hence, direct sampling methods preferable, but rarely possible for complicated densities in many dimensions.

Metropolis-Hastings Monte Carlo (cont)

A *Markov chain* is specified by two ingredients:
1) an initial state
2) a transition matrix $M(\vec{R}'|\vec{R})$ (probability of transition $\vec{R} \rightarrow \vec{R}'$.)

$$M(\vec{R}'|\vec{R}) \geq 0, \quad \sum_{\vec{R}'} M(\vec{R}'|\vec{R}) = 1. \qquad \textit{Stochastic matrix}$$

To sample $\rho(\mathbf{R})$, start from an arbitrary $\vec{R}$ and evolve the system by repeated application of M that satisfies the *stationarity condition*:

$$\sum_{\vec{R}'} M(\vec{R}'|\vec{R})\,\rho(\vec{R}) = \sum_{\vec{R}'} M(\vec{R}|\vec{R}')\,\rho(\vec{R}') = \rho(\vec{R}) \quad \forall\ \vec{R}$$

i.e., $\rho(\mathbf{R})$ is a right eigenvector of M with eigenvalue 1.
Stationarity $\Rightarrow$ if we start with ρ, will continue to sample ρ.
Want more than that: *any* initial density should evolve to ρ.

$$\lim_{n\to\infty} M^n(\vec{R}'|\vec{R})\,\delta(\vec{R}) = \rho(\vec{R}'), \quad \forall\ \vec{R}.$$

i.e., ρ should be the *dominant* right eigenvector.

Metropolis-Hastings Monte Carlo (cont)

Want that any initial density should evolve to ρ.

$$\lim_{n\to\infty} M^n(\vec{R}'|\vec{R})\delta(\vec{R}) = \rho(\vec{R}'), \quad \forall\ \vec{R}.$$

ρ should be the *dominant* right eigenvector.

In a finite space, necessary and sufficient condition on the Markov matrix M is that transitions can be made in a finite number of steps between any pair of states, that have nonzero probability. (*ergodic or irreducible matrix*.)

Same general idea holds in continuous space (matrix → integral kernel) but proofs are trickier.

In practice, length of Monte Carlo should be long enough that there be a significant probability of the system making several transitions between the neighborhoods of any pair of states that make a significant contribution to the average. This ensures that states are visited with the correct probability with only small statistical fluctuations. For example in a double-well system many transitions between the 2 wells should occur, but we can choose our proposal matrix to achieve this even if barrier between wells is high.

Metropolis-Hastings Monte Carlo (cont)

Construction of M

Impose *detailed balance* condition

$$M(\vec{R}'|\vec{R})\ \rho(\vec{R}) = M(\vec{R}|\vec{R}')\ \rho(\vec{R}')$$

Detailed balance more stringent than stationarity condition.
Detailed balance is not necessary but provides way to construct M.
Write elements of M as product of elements of a proposal matrix T and an acceptance Matrix A,

$$M(\vec{R}'|\vec{R}) = A(\vec{R}'|\vec{R})\ T(\vec{R}'|\vec{R})$$

$M(\vec{R}'|\vec{R})$ and $T(\vec{R}'|\vec{R})$ are stochastic matrices, but $A(\vec{R}'|\vec{R})$ is not.
Detailed balance is now:

$$A(\vec{R}'|\vec{R})\ T(\vec{R}'|\vec{R})\ \rho(\vec{R}) = A(\vec{R}|\vec{R}')\ T(\vec{R}|\vec{R}')\ \rho(\vec{R}')$$

$$\text{or} \qquad \frac{A(\vec{R}'|\vec{R})}{A(\vec{R}|\vec{R}')} = \frac{T(\vec{R}|\vec{R}')\ \rho(\vec{R}')}{T(\vec{R}'|\vec{R})\ \rho(\vec{R})}\ .$$

Metropolis-Hastings Monte Carlo (cont)

Choice of Acceptance Matrix A

$$\frac{A(\vec{R}'|\vec{R})}{A(\vec{R}|\vec{R}')} = \frac{T(\vec{R}|\vec{R}')\ \rho(\vec{R}')}{T(\vec{R}'|\vec{R})\ \rho(\vec{R})}\ .$$

Infinity of choices for A. Any function

$$F\left(\frac{T(\vec{R}|\vec{R}')\ \rho(\vec{R}')}{T(\vec{R}'|\vec{R})\ \rho(\vec{R})}\right) = A(\vec{R}'|\vec{R})$$

for which $F(x)/F(1/x) = x$ and $0 \leq F(x) \leq 1$ will do.
Choice of Metropolis *et al.* $F(x) = \min\{1, x\}$, maximizes the acceptance:

$$A(\vec{R}'|\vec{R}) = \min\left\{1, \frac{T(\vec{R}|\vec{R}')\ \rho(\vec{R}')}{T(\vec{R}'|\vec{R})\ \rho(\vec{R})}\right\}.$$

Other less good choices for $A(\vec{R}'|\vec{R})$ have been made, *e.g.* $F(x) = \frac{x}{1+x}$

$$A(\vec{R}'|\vec{R}) = \frac{T(\vec{R}|\vec{R}')\ \rho(\vec{R}')}{T(\vec{R}|\vec{R}')\ \rho(\vec{R}') + T(\vec{R}'|\vec{R})\ \rho(\vec{R})}.$$

Metropolis: $T(\vec{R}|\vec{R}') = T(\vec{R}'|\vec{R})$, Hastings: $T(\vec{R}|\vec{R}') \neq T(\vec{R}'|\vec{R})$

Metropolis-Hastings Monte Carlo (cont)

Choice of Proposal Matrix T

So, the optimal choice for the acceptance matrix $A(\vec{R}'|\vec{R})$ is simple and known.

However, there is considerable scope for using one's ingenuity to come up with good proposal matrices, $T(\vec{R}'|\vec{R})$, that allow one to make large moves with large acceptances, in order to make the autocorrelation time is small.

Metropolis-Hastings Monte Carlo (cont)

Some Observations about Metropolis Method

1. To sample states with relative density ρ it is not necessary to know the normalization of ρ. Metropolis automatically samples $\rho / \sum \rho$.
2. The variance of the estimate for the expectation value $\langle X \rangle$ is given by

$$\frac{T_{\rm corr}}{N-1} \left(\frac{\sum X(\vec{R})^2}{N} - \left(\frac{\sum X(\vec{R})}{N} \right)^2 \right).$$

   That is, the effective number of configurations $N_{\rm eff}$ is smaller than N by a factor of $T_{\rm corr}$, which we define to be the autocorrelation time. ($T_{\rm corr}$ is related to integrated autocorrelation time, $T_{\rm corr} = 1 + 2t_{\rm corr}$.)

3. The rate of convergence to the desired density and the autocorrelation time of estimates of observables is governed by the sub-dominant eigenvalues of M. In practice reduce $T_{\rm corr}$ by inventing large moves that have large acceptance probabilities.

Metropolis-Hastings Monte Carlo (cont)

Some Observations about Metropolis Method

4. Folklore: when one can choose from a range of proposal matrices, the optimal one has an average acceptance ratio close to $1/2$.
   In fact the optimal choice may have an average acceptance that is anywhere between zero and one.
   I have found instances where the optimum is as small as 0.2 or as large as 0.9.
   A much better criterion is to maximize the rate at which the system diffuses through configuration space $\langle A(\vec{R}'|\vec{R})(\vec{R}'-\vec{R})^2\rangle$.
   The real measure of goodness is of course to minimize the autocorrelation time for the observables of interest.

Metropolis-Hastings Monte Carlo (cont)

Some Observations about Metropolis Method

5. If $M_1, M_2, \cdots, M_n$ are Markov matrices that satisfy the stationarity condition.
   Combine these to construct a compound transition matrix M that also satisfies the stationarity condition. One often employs elementary transition matrices M_i that are non-ergodic to construct a compound transition matrices M that is ergodic.
   Two sorts of combinations are often useful:
   1. $M = \Pi_{i=1}^{n} M_i$. Sequential updating. e.g. Ising spins on lattice sites or the electrons in electronic structure.
   2. $M = \sum_{i=1}^{n} c_i M_i$, $c_i \geq 0$ (independent of the current state) and $\sum_{i=1}^{n} c_i = 1$. Choose the transitions M_i randomly with probabilities c_i.

Metropolis-Hastings Monte Carlo (cont)

Some Observations about Metropolis Method

7. The *heat bath* algorithm can be considered to be a special case of the (generalized) Metropolis method. $T(\vec{R}'|\vec{R}) \propto \rho(\vec{R}')$ for only a small set of accessible states in a domain $D(\vec{R})$ in the neighborhood of $\vec{R}$:

$$T(\vec{R}'|\vec{R}) = \begin{cases} \rho(\vec{R}')/\sum\rho(\vec{R}') & \text{if } \vec{R}' \epsilon\, D(\vec{R}) \\ 0 & \text{otherwise} \end{cases}$$

If the sum over the accessible states from $\vec{R}$ and $\vec{R}'$ is the same, the acceptance is unity. The heat bath algorithm frequently used for lattice models where the normalization constant can be easily computed, e.g. Potts model.

Metropolis-Hastings Monte Carlo (cont)

Autocorrelation time

N Monte Carlo steps $=$ N_b blocks $\times$ N_s steps/block
N measurements of $E_{\rm L}$

$$\begin{aligned}\bar{E} &= \text{average of } E_{\rm L} \\ \sigma &= \text{rms fluctuations of individual } E_{\rm L} \\ \sigma_b &= \text{rms fluctuations of block averages of } E_{\rm L}\end{aligned}$$

Effectively, $N/T_{\rm corr}$ independent measurements of $E_{\rm L}$

Define $T_{\rm corr}$ as

$$\mathrm{err}(\bar{E}) = \frac{\sigma}{\sqrt{N_b \times N_s}} \sqrt{T_{\rm corr}} = \frac{\sigma_b}{\sqrt{N_b}}$$

$$\Rightarrow \quad \boxed{T_{\rm corr} = N_s \left(\frac{\sigma_b}{\sigma}\right)^2}$$

Choose $N_s \gg T_{\rm corr}$, say, 100 $T_{\rm corr}$.
If $N_s \approx 10\, T_{\rm corr}$, $T_{\rm corr}$ underest. $\approx$ 10%.