Monte Carlo Methods, II

M. D. Jones, Ph.D.

Center for Computational Research
University at Buffalo
State University of New York

High Performance Computing I, 2013

Markov Processes

Working definitions of **stochastic** and **Markov** processes:

- A **stochastic process** is a movement through well-defined states using some form of random selection.
- A **Markov process** is a stochastic process that has no memory of prior states, namely one whose next state depends only on the current state. Described by a set of **transition probabilities**.
- A random walk on a discrete grid is a good example of a Markov chain (when the walk is unconstrained).

Monte Carlo Integration

- Consider a concrete example, the definite integral

$$I = \int_0^1 g(x)dx,$$

- Terminology:
  - **Mean**, $\langle I \rangle$,

$$\langle I \rangle = \frac{1}{N}\sum_{i=1}^{N} S_i = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M} g(x_j)$$

More MC Integration

More Monte Carlo quadrature terminology (cont'd):

- **Variance**, σ^2,

$$\sigma_g^2 = \langle g^2 \rangle - \langle g \rangle^2,$$

$$\sigma_g^2 = \frac{1}{N} \sum_{i=1}^{N} g_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} g_i \right)^2,$$

- **Error**, σ_I,

$$\sigma_I^2 = \sigma_g^2 / (N - 1).$$

MC vs. Other Quadrature Methods

Compare MC with other numerical quadrature techniques:

- MC converges slooooowly, only as $1/\sqrt{N}$,

$$\sigma_I = \sqrt{\sigma_g^2/(N-1)}.$$

- Compare and contrast with, say, the trapezoidal rule

$$\int_{-h}^{h} g(x)dx \sim \frac{h}{2}\left[g(-h) + 2g(0) + g(h)\right] + \mathcal{O}(h^3),$$

which, for N sample points, $h \sim 1/N$, so the trapezoidal rule converges like $1/N^3$.

The Monte Carlo Way

- So if the Trapezoidal rule converges like $1/N^3$, why use MC?
- Consider now N points in d dimensions. Each dimension gets $N^{1/d}$ intervals, or $h \sim N^{-1/d}$. The trapezoidal rule in each dimension leads to errors $\mathcal{O}(h^{d+2})$. For all N cells the total error comes to

$$N\mathcal{O}(h^{d+2}) \sim N\left[\left(N^{-1/d}\right)^{d+2}\right] = N^{-2/d},$$

So for $d \geq 4$ the advantage goes to MC. And there are **many** cases (not just thermodynamic, like the partition function) where $d >> 4$.

Monte Carlo Justified

- Monte Carlo is an effective technique for high dimensional integration.
- Consider the following integral,

$$\mathcal{Z} = \int d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_{N_p} e^{-\beta \sum_{i<j} v(r_{ij})},$$

  which is the classical partition function for a gas of N_p atoms at (inverse) temperature β interacting through a pair potential $v(r_{ij})$.
- This example is a $3N_p$ dimensional integral.

Simple MC Example

Let us consider a simple example - our old friend:

- Evaluate π using the "dart-throwing" method,

$$\frac{\pi}{4} = \int_0^1 \int_0^1 g(x,y)dxdy,$$

$$g(x,y) = \begin{cases} 1 & x^2 + y^2 \leq 1, \\ 0 & x^2 + y^2 \geq 1. \end{cases}$$

- Trivial example, but quite useful for purpose of demonstration.

Simple MC Example (cont'd)

In sequential code we would have something like the following:

```
1           sum_g=0.0;
2           for(i=1;i<=N;i++) {
3               x=URN();
4               y=URN();
5               if (x*x+y*y <= 1.0) sum_g += 1.0;
6           }
7           sum_g /= N;
8           sig_g = SQRT( (sum_g-sum_g*sum_g)/(N-1) );
9           printf(``pi/4 = %f +- %f\n'',sum_f,sig_f);
```

Simple MC Example (cont'd)

In parallel: (using independent RN streams)

```
1  iseed=123+10*myRank;
2  init_URN(iseed);
3  my_share=N/numProcs;
4  my_sum=0.0;
5  for(i=1;i<=my_share;i++) {
6       x=URN(iseed);
7       y=URN(iseed);
8       if (x*x+y*y <= 1.0) { my_sum += 1.0; }
9  }
10 MPI_Reduce(&my_sum,&sum_f,1,MPI_DOUBLE,MPI_SUM,0,MPI_COMM_WORLD);
11 if (myRank == 0) {
12     sum_g /= N;
13     sig_g = SQRT( (sum_g-sum_g*sum_g)/(N-1) );
14     printf(``pi/4 = %f +- %f\n'',sum_g,sig_g);
15 }
```

Simple MC Example (cont'd)

- Consider the variance - can you work it out analytically?

$$\sigma_g^2 = \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 = 0.168$$

- To reduce error to 10^{-6} therefore requires almost 2×10^{11} steps
- What can we do to reduce the "intrinsic" variance?
- Answer - much of the Monte Carlo literature is devoted to techniques for **variance reduction**

Law of Large Numbers

- Requires only that variance exists (can still be formulated for other cases, but harder)
- "Law of Large Numbers," given independent (**uncorrelated!**) RNs $\{x_1, \ldots, x_N\}$ each has expectation value μ and variance σ^2.
- As $N \to \infty$,

$$P\{\lim_{N\to\infty} \bar{x}_N = \mu\} = 1$$

  where $\bar{x}_N = \frac{1}{N}\sum_{i=1}^{N} x_i$.
- So, in essence, this "law" implies that the mean of N random samples converges (probabilistically) to its expected value.
- Now we can say something a bit stronger about the rate of convergence ...

Chebychev Inequality

Given estimator G, mean $\langle G \rangle$, variance $\mathtt{var}\{G\}$, the Chebychev inequality is:

$$P\left\{ |G - \langle G \rangle| \geq \left[\frac{\mathtt{var}\{G\}}{\delta} \right]^{1/2} \right\} \leq \delta,$$

where $\delta > 0$.

- Basically an estimate for getting a large deviation in a MC calculation
- As $N \to \infty$, and $\mathtt{var}\{G\} = \mathtt{var}\{g\}/N$, probability of large deviation becomes small
- Can be made stronger still using central limit theorem ...

Central Limit Theorem

In a nutshell, the central limit theorem states that:

Central Limit Theorem

If the sum of many independent identically distributed random variables has a finite variance, then it will tend to be approximately normally distributed

- Many natural processes are derived from distributions with a finite variance, hence are normally distributed
- E.g., human height

- Given estimator for N samples,

$$G_N = \frac{1}{N}\sum_{i=1}^{N} g(x_n)$$

- Let t_N denote the deviation from the expectation,

$$\begin{aligned} t_N &= \left(G_N - \langle g \rangle\right) / \left[\mathtt{var}\{G_N\}\right]^{1/2}, \\ &= \frac{\sqrt{N}(G_N - \langle g \rangle)}{\left[\mathtt{var}\{G_N\}\right]^{1/2}} \end{aligned}$$

- PDF,

$$f(G_N) = \frac{1}{\sqrt{2\pi(\sigma^2/N)}} \exp\left[-\frac{N(G_N - \langle g \rangle)^2}{2\sigma^2}\right]$$

  i.e. the observed G_N is within one standard error ($\sigma/\sqrt{N}$) 68.3% of the time, two standard errors 95.4%, etc.

Implications of the Central Limit Theorem

The central limit theorem has several implications:

- Applies asymptotically, satisfied when third central moment satisfies:

$$|\mu_3| \ll \sigma^3 \sqrt{N}$$

- In a real MC simulation, can (**should!**) monitor distribution of estimators
- Many MC simulations **assume** central limit theorem is satisfied, no matter the size of N - reported errors are likely to be wrong
- Infinite variance - sometimes still possible to find limit distribution, likely not normal

Importance Sampling

Importance sampling:

- Importance Sampling: Recast the integral,

$$I = \int f(\mathbf{R}) d\mathbf{R} = \int p(\mathbf{R}) \frac{f(\mathbf{R})}{p(\mathbf{R})} d\mathbf{R},$$

- and sample according to the probability distribution function (PDF) $p(\mathbf{R})$,

$$\langle f \rangle_p = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{R})}{p(\mathbf{R})}.$$

- The only "trick" is to choose a PDF $p(\mathbf{R})$ that closely mimics the original integrand $f(\mathbf{R})$.

Importance Sampling

- New variance expression,

$$\sigma^2 = \int d\mathbf{R} p(\mathbf{R}) \left(\frac{f}{p} - I \right)^2 = \int d\mathbf{R} \left(\frac{f^2}{p} - I \right).$$

- Optimal importance function, p^*,

$$p^*(\mathbf{R}) = \frac{|f(\mathbf{R})|}{\int d\mathbf{R} |f(\mathbf{R})|}.$$

- Zero-variance Principle: when f is entirely of one-sign, an arbitrarily good choice can be made for p.

Importance Sampling Example

- A simple one-dimensional integral,

$$I = \int_0^L \frac{e^{-x^2}}{1 + 10x^2} dx,$$

  the integrand decays very quickly - uniform sampling will not be efficient.
- Introduce importance function $p(x) = e^{-ax}$, where a is a tunable parameter.

Monte Carlo Efficiency

- Comparing variances is not good enough (sampling a function $p(x)$ is more computationally demanding) - introduce the efficiency, ξ,

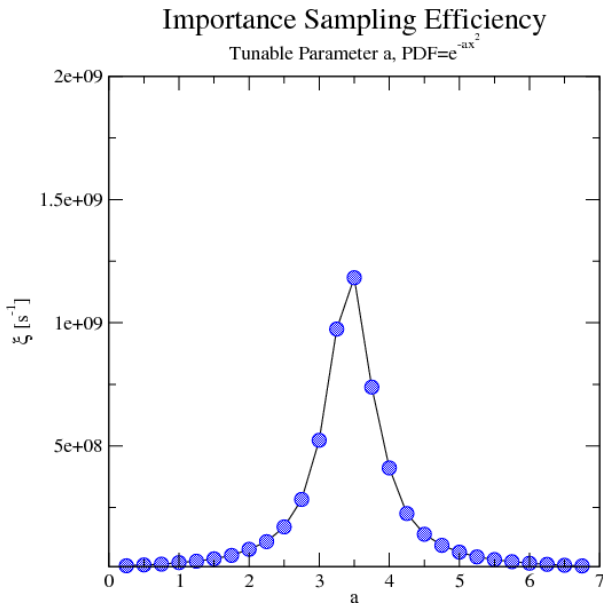
$$\xi = \frac{1}{\tau \sigma_I^2},$$

where τ is the time required to achieve a variance σ_I^2 in the estimate of the integral I.

IS Example Results

Back to our example problem, with tunable parameter *a* in the PDF.

- Tuning the parameter *a*,



Note that the baseline for uniform sampling is an efficieny of only $\sim 10^7$.

IS Example Results (cont'd)

- Importance sampling MC efficiency compared directly with uniform sampling, for a range of L.

L	IS(e^{-ax})	US(1)
1.0	1.2×10^9	3.0×10^7
10.0	8.5×10^8	1.4×10^6
100.0	8.5×10^8	9.1×10^4

- Importance sampling is critical when the integrand is highly non-uniform.

Metropolis Algorithm

- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
- a.k.a. *Markov Chain Monte Carlo*.
- General method for generating a sequence of configurations via a random walk $\{\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_n\}$ that are distributed according to probability distribution p.
- $T(\mathbf{R} \rightarrow \mathbf{R}')$, probability of making a trial move from configuration $\mathbf{R}$ to $\mathbf{R}'$.
- Detailed Balance:

$$p(\mathbf{R})T(\mathbf{R} \rightarrow \mathbf{R}') = p(\mathbf{R}')T(\mathbf{R}' \rightarrow \mathbf{R}).$$

Metropolis Recipe

- Generalized metropolis sampling rule: accept trial step with probability

$$a = \min(1, q),$$
$$q = \frac{T(\mathbf{R}' \rightarrow \mathbf{R})p(\mathbf{R}')}{T(\mathbf{R} \rightarrow \mathbf{R}')p(\mathbf{R})}.$$

- only requirement is that we can evaluate $p(\mathbf{R})$ for a given configuration.
- Not limited to continuous variables - any combination of discrete and spatial variables will work.

Metropolis Example

- Evaluation of thermodynamic integrals,

$$\mathcal{Z} = \int d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_{N_p} e^{-\beta \sum_{i<j} v(r_{ij})},$$

- Consider sampling classical Boltzmann distribution, $e^{-\beta V(\mathbf{R})}$. (logical choice - can you imagine sampling this integral uniformly?)

Metropolis Example (cont'd)

- Choose $T(\mathbf{R} \rightarrow \mathbf{R}')$ to be constant inside a cube of side length δ, zero outside the cube,

$$\mathbf{R}' = \mathbf{R} + (\texttt{URN}() - 0.5)\,\delta,$$

- Acceptance is then based on $e^{-\beta[V(\mathbf{R}')-V(\mathbf{R})]}$.
- Trial moves that increase the potential are acceptable, if the thermal energy cost is relatively small.

Metropolis Example 2 (2D Ising Model)

Consider a 2D spin system on a $L \times L$ lattice of points, $s_{ij} = \pm 1$ are the individual spin values. Total energy is given by each spin interacting only with its nearest neighbor (4 in 2D):

$$E[\{s_{ij}\}] = -\frac{1}{2}\sum_{ij} s_{ij}(s_{l(ij)} + s_{r(ij)} + s_{u(ij)} + s_{d(ij)}),$$

where the subscripts $l/r/u/d(ij)$ denotes the left/right/up/down neighbor of spin (i,j). The Boltzmann factor, W, determines the probability of finding the system in a particular spin configuration, $\{s_{ij}\}$,

$$W[\{s_{ij}\}] = \exp\left(-\beta E[\{s_{ij}\}]\right)/Z,$$

and Z is the partition function, $Z = \sum_{s_{ij}=\pm 1} \exp\left(-\beta E[\{s_{ij}\}]\right)$, summed over all possible spin states ($\beta = (k_B T)^{-1}$).

The average magnetization, $\langle M \rangle$, is given by averaging over all possible spin settings, counting each with its probability weight, W:

$$\langle M \rangle = \sum_{s'_{ij}=\pm 1} \left(W[\{s'_{ij}\}] \sum_{ij} s'_{ij} \right)$$

Not a summation that you are going to easily evaluate analytically (although L. Onsager solved it in 2D, c.f. Phys. Rev. **65**, 117 (1944).), but it is quite amenable to the Metropolis Monte Carlo approach.

Use the Metropolis formalism to generate a sequence of randomly generated spin configurations with the probability distribution $W[\{s_{ij}\}]$:

1. Start with randomly generated spin configuration, $\{s_{ij}\}$
2. Choose point (i, j), trial step - flip its spin with energy change ΔE. If $\Delta E \leq 0$, accept change, otherwise accept with probability $P(\Delta E) = \exp(-\beta \Delta E)$ (i.e. accept the trial flip if URN < $P(\Delta E)$).
3. Repeat previous step for all remaining lattice sites (sweep lattice)
4. Metropolis MC estimate for $\langle M \rangle$,

$$\langle M \rangle = \frac{1}{N} \sum_{k=1}^{N} \sum_{ij} s_{ij}^{k},$$

where $\{s_{ij}^{k}\}$ is the sequence of spin configurations so generated, typically "thermalized" over a certain number of initial sweeps.

Metropolis Rules of Thumb

Some general rules of thumb when using the Metropolis algorithm:

- Acceptance ratio: generally close to 0.5. You still have to look at efficiency.
- Flexible trial steps: particles can be moved one at a time. Compare this with MD. One can always invent better transition rules.
- Optimal: one can show that the Metropolis algorithm is optimal compared to similar schemes.

Smarter Monte Carlo

A profusion of schemes (and associated acronyms) for “smarter” Monte Carlo:

- Basically all of them are schemes for variance reduction
- Improved sampling methods, generally using Metropolis

Force-Bias Monte Carlo

- Motivation: particle movements biased in direction of forces acting on them
- M. Pangali, C. Rao, and B. J. Berne, Mol. Phys **37**, 1773-1798 (1979).
- single-particle i moves from state $m \rightarrow n$,

$$\alpha_{mn} \sim \exp\left[\lambda\beta(\mathbf{f}_i^m \cdot \delta\mathbf{r}_i^{nm})\right]$$

where $\delta\mathbf{r}_i^{nm} = \mathbf{r}_i^n - \mathbf{r}_i^m$, λ is a constant

Reblocking for Better Averages

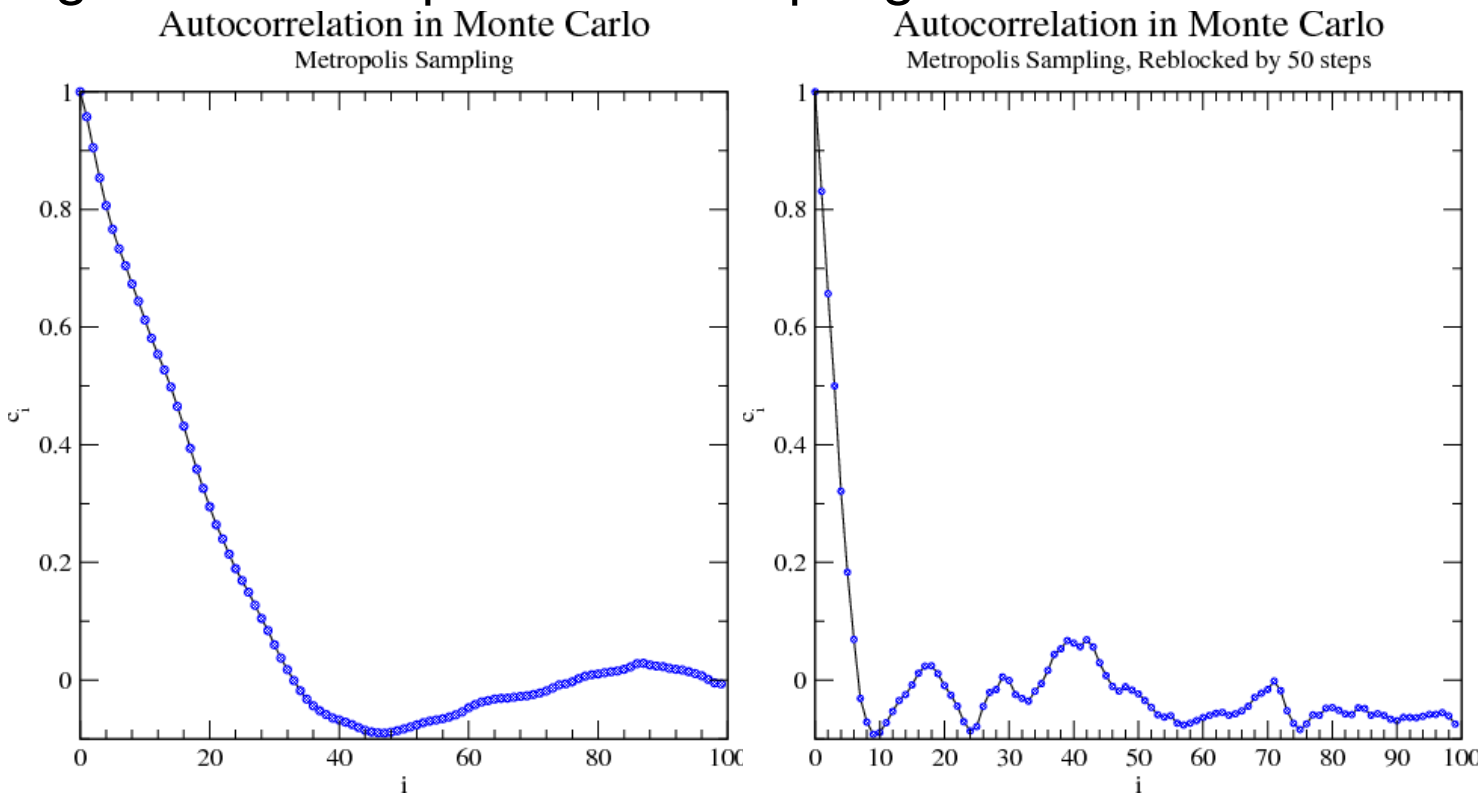
- It is often the case that successive MC steps are strongly correlated, which can be found by studying the autocorrelation coefficient,

$$c_i = \frac{1}{(N-i)\sigma_f^2} \sum_{j=1}^{M-i} (f_j - \langle f \rangle)(f_{j+i} - \langle f \rangle),$$

- In practice, one studies c_i to determine the optimal reblocking interval (when $c_i \sim 0.1$).

Autocorrelation Example

- Practical example - a Monte Carlo example that uses Metropolis algorithm for importance sampling.



Quantum Monte Carlo

Quantum Monte Carlo (QMC) methods are a specialized class of methods for the solution of quantum many-body problems. Given a Hamiltonian operator, $\hat{H}$, and a trial wavefunction, $\Psi(\mathbf{r})$, where $\mathbf{r}$ is a $3N$-dimensional coordinate vector,

$$E = \int d\mathbf{r} \left(\frac{|\Psi(\mathbf{r})|^2}{\int d\mathbf{r} |\Psi(\mathbf{r})|^2} \right) \left[\Psi^{-1}(\mathbf{r}) \hat{H} \Psi(\mathbf{r}) \right]$$

is the energy estimator. The trial wavefunction need not even be normalized, and can be sampled using Metropolis.

Variations of QMC

QMC comes in several flavors:

- **Variational Monte Carlo** (VMC) - relies on good approximate parameterization of wave function
- **Diffusion Monte Carlo** (DMC) - solve Schrödinger equation in imaginary time
- **Green's Function Monte Carlo** (GFMC) - stochastic propagation using natural Green's function formalism
- **Path Integral Monte Carlo** (PIMC) - finite temperature formulation using Feynman path integrals

Financial Application of MC

MC is also widely used in financial modeling - in 1997 Black and Scholes received the Nobel prize in Economics for their theory of treating valuations (e.g. stock price) as Brownian motion:

$$dS = \mu S dt + \sigma S dW_t,$$

where S denotes value, μ the (annualized) drift rate, σ the volatility, and W_t the volatility/risk history (Wiener process), i.e. any and all known sources of uncertainty in the price history of S.
Typically sample a path from 0 to T in M units of δt:

$$S(k\delta t) = S(0) \exp\left(\sum_{i=1}^{k}\left[\left(\mu - \sigma^2/2\right)\delta t + \sigma\epsilon_i\sqrt{\delta t}\right]\right),$$

where ϵ_i is a Gaussian random variate.
MC is handy when there are many sources of uncertainty, and hence many potential degrees of freedom, so not as straightforward as above.

More Information

Monte Carlo is very widely applied - some good general references:

- Hammersley, J. M. and D. C. Handscomb, *Monte Carlo Methods*, (John Wiley & Sons, New York, 1964).
- M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods, Volume I: Basics* (John Wiley & Sons, New York, 1986).

... and a couple of good review articles (albeit rather specialized):

- Foulkes, W. M. C. and Mitas, L. and Needs, R. J. and Rajagopal, G., *Quantum Monte Carlo simulations of solids*, Rev. Mod. Phys. **73**, 33-83 (2001).
- Jensen, P., *Growth of nanostructures by cluster deposition: Experiments and simple models*, Rev. Mod. Phys. **71**, 1695-1735 (1999).