Atomic Scale Simulation

Introduction to Monte Carlo

What is Monte Carlo?

Monte Carlo is a way of solving problems using "random numbers" in some essential way. All Monte Carlo calculations can be viewed as a multidimensional integration, although that may not always be the most useful point of view. Consider the following multidimensional integral:

$$I = \int_0^1 d^D x f(x). \tag{1}$$

With MC, we sample randomly in the "cube" and take the average value.

$$I = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f(x_i). \tag{2}$$

We will see in a minute that the (statistical) error ϵ of our estimate of I converges like $\epsilon \propto M^{-1/2}$. Hence, the computer time T, needed to get an accuracy ϵ goes like

$$T_{MC} \propto \epsilon^{-2}. \tag{3}$$

One order of magnitude in accuracy requires a 100 times longer run! This is true in all stochastic or Monte Carlo methods.

Why do we then use Monte Carlo? Consider the alternatives. Suppose we use the trapezoidal rule. There we find the error is $\epsilon \propto dx^2 f''(x)$. The improved Simpson's rule gives an error $\epsilon \propto dx^4 f^{(4)}(x)$ where $f^{(4)}$ is the fourth derivative and dx is the grid spacing. In general we can write that the error is proportional to:

$$\epsilon \propto dx^\alpha c_\alpha \tag{4}$$

for an integration rule good to order α where c_α is proportional to the maximum α^{th} derivative of the function in the integration region. But here is the problem: the computer time will scale with the number of points: dx^{-D}. Hence we find that the time to do a "traditional" integration goes as:

$$T_{int} \propto \epsilon^{-D/\alpha} \tag{5}$$

Hence in the limit of small ϵ, Monte Carlo will be more efficient if $D > 2\alpha$.

Why is this? It simply takes too long to fill an D-dimensional space uniformly with points. To get a reasonable value for the integrand, the grid spacing dx must be smaller than the natural variation in $f(x)$. Typically hundreds of grid points in each dimension are required. If one raises one hundred to some high power, one cannot complete the integration.

Why can't we use very high order integration schemes, that is, take α large? The order of the scheme α is controlled by the smoothness of the function. Typical functions that we use in science become nasty if you differentiate, them too often. This means that higher order schemes

are *less* accurate than low order schemes unless dx is exceedingly small. Hence one gets best results for $2 \leq \alpha \leq 4$. That means that MC is preferred once you get above 6 dimensions or so. This is borne out by experience.

The other reasons for doing MC are its conceptual simplicity and the fact that it comes with built in error bars. But the scaling behavior is the crucial difference.

Later we will discuss an alternative integration method which is between MC and grid based methods called "quasi-random numbers."

Monte Carlo Terminology

First let us review some terms from probability theory.

A *probability distribution function* $p(x)dx$ is the probability of x being in a small interval $(x, x+dx)$. This means than $p(x) \geq 0$ and $\int dxp(x) = 1$. I will not particularly worry about what x is, a real number or an integer (e.g. the state of a thrown die.) The *cumulative distribution* is better defined mathematically. $F(x)$ is the probability that $y \leq x$. Hence $F(x) = \int_{-\infty}^{x} dy\, p(y)$.

The *expectation value* of some function $f(x)$ with respect to p is $I =< f >= \bar{f} = \int dx f(x)p(x)$. With an *estimator* we *sample* a set of N values $\{x_i\}$ from p and use those to estimate the value of I. The best estimator for the simple integral above is:

$$f_N = \frac{\sum_{i=1}^{N} f(x_i)}{N} \tag{6}$$

Now f_N is distributed according to some probability distribution p_{f_N}. Of course, by construction $< f_N >= I$. But we can ask, what are the fluctuations of f_N about its mean value? This is the *variance* of f_N.

The *central limit theorem* is the foundation on which MC is built. The estimator, Eq.(6) will converge to the value of the integral, if the variance of $f(x)$ exists. Furthermore, we will know exactly the probability distribution of f_N in the limit of large N; it is the normal distribution:

$$p(f_N) = (2\pi\sigma_N^2)^{-1/2} \exp[-(f_N - I)^2/(2\sigma_N^2)] \tag{7}$$

Since the successive values of x_i are uncorrelated we have the variance of the mean is

$$\sigma_N^2 = \frac{1}{N}\sigma^2 = \frac{1}{N}\langle [f(x) - I]^2 \rangle = \frac{1}{N}\int dxp(x)[f(x) - I]^2 \tag{8}$$

Traditionally in MC, one quotes 1 σ deviations as the error. Since we know it is a normal distribution, this means that 33% of the time one expects the value to exceed 1 σ, 5%, a 2 σ deviation etc. A 2 σ error should not concern you too much, it happens fairly often.

Of crucial importance in any MC work is to prove (analytically) that the variance σ exists. Only then you can talk about reliably about estimating errors. Usually the variance (of f_N) is estimated itself from the data with the formula:

$$\sigma^2 = \int dx(f(x) - I)^2 = \frac{1}{N-1}\sum_{i=1}^{N}[f(x_i) - f_N]^2 \tag{9}$$

with the F_N estimated from eq. (6). The $N-1$ appears because both the mean and variance are estimated from the same data. This formula is only correct if the data points are uncorrelated; otherwise one has to do blocking to eliminate the correlation. Then the estimate of the error of the mean is $\sigma_N = \sqrt{\sigma^2/N}$.

Multidimensional distributions

Up to now we have been discussing a single variable (x or f). Often the output of a MC simulation is several variables (e.g. the energy and the pressure). We will write the M output variables as a vector $\mathbf{x} = (x_1, \ldots, x_M)$. Then there is a p.d.f. in that M-dimensional space, $p(\mathbf{x})d\mathbf{x}$. One can generalize the central limit theorem for that case. Define the *covariance matrix*:

$$\nu_{i,j} = \langle (x_i - < x_i >)(x_j - < x_j >) \rangle \tag{10}$$

The covariance matrix is a positive symmetric matrix. This means that it can be diagonalized and has positive eigenvalues.

Assume $p(\mathbf{x})d\mathbf{x}$ is such that all elements of $\nu_{i,j}$ are finite, the various variances exist. Then the central limit theorem says that the estimate of the mean:

$$\mathbf{x}_N = \frac{\sum_{k=1}^{N} \mathbf{x}_k}{N} \tag{11}$$

for sufficiently large N will obey the multidimensional normal (Gaussian) distribution:

$$p_N(\mathbf{x}_N) = [2\pi det(\nu)/N]^{-1/2} \exp[-(\mathbf{x}_N - < \mathbf{x}_N >)\frac{N\nu^{-1}}{2}(\mathbf{x}_N - < \mathbf{x}_N >)] \tag{12}$$

where the matrix inverse and determinant are meant. Unless ν is diagonal, the various values of ν are *correlated* positively if $\nu_{i,j} > 0$, negatively otherwise. The covariance matrix can be estimated from applying its definition (Eq. (10) to a sequence of $\mathbf{x}_i$'s. One note of caution: it takes at least $(M+3)/2$ vectors of data to fully specify the covariance matrix. Many more usually are needed.

A & T pgs. 110-118
Kalos-Whitlock pgs. 1-38
Hammersley-Handscomb pgs. 1-24