ERROR ANALYSIS IN MONTE CARLO

This note is intended to review and clarify the comments I made about error analysis in Monte Carlo, and illustrate them with the very simple example of a simulation of $E(x) = \frac{1}{2}kx^2$.

]

I. INTRODUCTION AND DEFINITIONS

We begin with the definitions,

$$\langle x \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\langle x^2 \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \tag{2}$$

$$\sigma = \sqrt{\frac{\langle x^2 \rangle - \langle x \rangle^2}{N-1}}. \tag{3}$$

Here x_i is the ith measured value of x, and N is the number of measurements. The definitions of $\langle x \rangle$ and $\langle x^2 \rangle$ are unambiguous. The entire content of this note is to clarify the proper denominator of the definition of σ. Specifically, the formula for σ assumes that the N measurements are all independent. Since successive values of x are generated from each other, this is never true. The x values are more and more related the less time one waits between measurements.

To quantify the correlations among successive x values, we define the autocorrelation function,

$$c(l) = \frac{1}{N-l}\sum_{i=1}^{N-l} y_i y_{i+l} \tag{4}$$

Here $y(i) = x(i) - \langle x \rangle$ measures the fluctuation of x about its average value. $c(l)$ measures whether those fluctuations are related for x values l measurements apart. Saying x_i and x_{i+l} are independent means whether x_{i+l} is above or below $\langle x \rangle$ (the sign of y_{i+l}) is unrelated to the sign of y_i. If that is the case, $c(l) = 0$ (to within errors). Clearly $c(0)$ is never zero. In fact, $c(0) = \sigma^2$. It is conventional to redefine $c(0) \to c(0)/\sigma^2$ so that $c(0) = 1$.

II. TIME HISTORIES

Let's begin by looking at actual time histories of x. I chose $k = 1$ and $T = 2$ so that $\langle x^2 \rangle = 2$. The step size for suggested changes in x is Δ. I measure every monte carlo step and ran for $N = 400000$ sweeps. Here of course since $\langle x \rangle = 0$, $y_i = x_i$. The Monte Carlo time histories are given in Figure 1. That the data are correlated is immediately evident. If a value of x_i is positive, its successors tend to be positive and similarly if it is negative. The dependence on the step size Δ is easy to interpret.

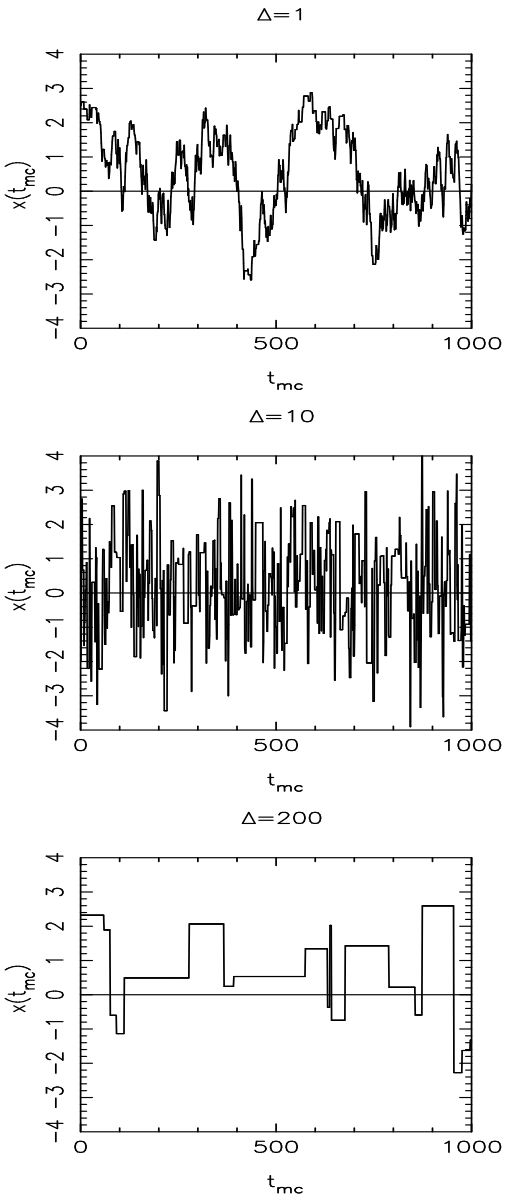

FIG. 1: First 1000 steps in Monte Carlo time history of x for three different step sizes $\Delta = 1, 10, 200$. (Acceptance rates=0.70, 0.35, 0.02).

If Δ is small you do not suggest much of a change, and successive values of x are highly correlated. Likewise, if Δ is large, most suggested Monte Carlo moves take you out of the part of phase space of low energy and are rejected. (This results in the long flat regions of the evolution of x.)

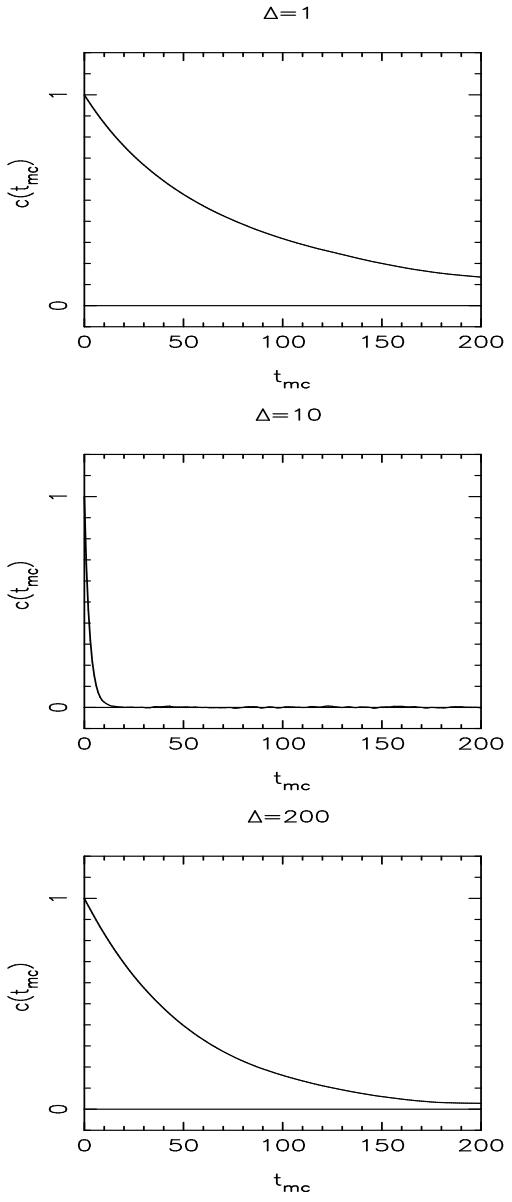

FIG. 2: Autocorrelation functions for the complete data sets (400,000 steps) as in figure 1. (Step sizes $\Delta = 1, 10, 200$, Acceptance rates=0.70, 0.35, 0.02).

III. CORRELATION FUNCTIONS

The plots of $c(l)$ resulting from the same data are given in Figure 2. We see that $c(l)$ has a characteristic decaying exponential form. We define the correlation time τ to be the point when $c(l = \tau) = e^{-1}$ and say that measurements of x are independent when l exceeds τ. (Strictly speaking, we want c to go to zero, but $c(\tau) = e^{-1}$ is an acceptable criterion.) Notice you can almost guess the values of τ given by Figure 2 directly from the time histories of Figure 1.

As mentioned earlier, in generating the above results I measured x every Monte Carlo step. What happens if one instead measures only every mth step? Define $c_m(l)$ to be the correlation function when measurements are only every mth Monte Carlo step. It is easy to convince yourself that $c_m(l) = c_1(ml)$, so the correlation function rescales in a trivially fashion. The point is that if one choose $m > \tau$, then the measurements all become independent.

• *So one way to ensure the value for the error bar σ is correct is to make sure measurements are separated by a waiting time $m > \tau$.*

This approach has the advantage that one does not waste time making measurements when the measurements are not independent.

IV. REBINNING DATA

An alternate (and equivalent) approach to getting the correct σ is by "rebinning" the data. Take a file containing the complete time history of a measurement, for example the data for x which is partially shown in Figure 1. Choose a "bin size" M, and average the data for x over each of the $L = N/M$ bins (remember N = total number of measurements) to create L "binned measurements" m_j.

$$m_j = \frac{1}{M} \sum_{i=M*(j-1)+1}^{M*j} x_i. \tag{5}$$

Treat these L values for m as your independent measurements. As seen in Equation 5, the values for m are already averages over M values of x. Define averages and error bars as in Equation 1, with L replacing N in the normalizations $1/N$ and $1/\sqrt{N-1}$. The average $\langle x \rangle$ is independent of M since all one is doing is reordering a linear sum. The average $\langle x^2 \rangle$ is however altered, and hence so is the error bar σ. Figure 3 shows values for σ as a function of the number of x values in each bin, M.

What is the interpretation of Figure 3? Consider $M = 1$. In this case only one value of x is put in each bin, so that in calculating σ it is assumed all x are independent. The error bar σ thus obtained is too small. As M becomes larger, more x values are put in each bin, the number of bins (independent measurements) decreases, and σ increases. Eventually σ goes to an asymptotic value which gives the correct error bar.

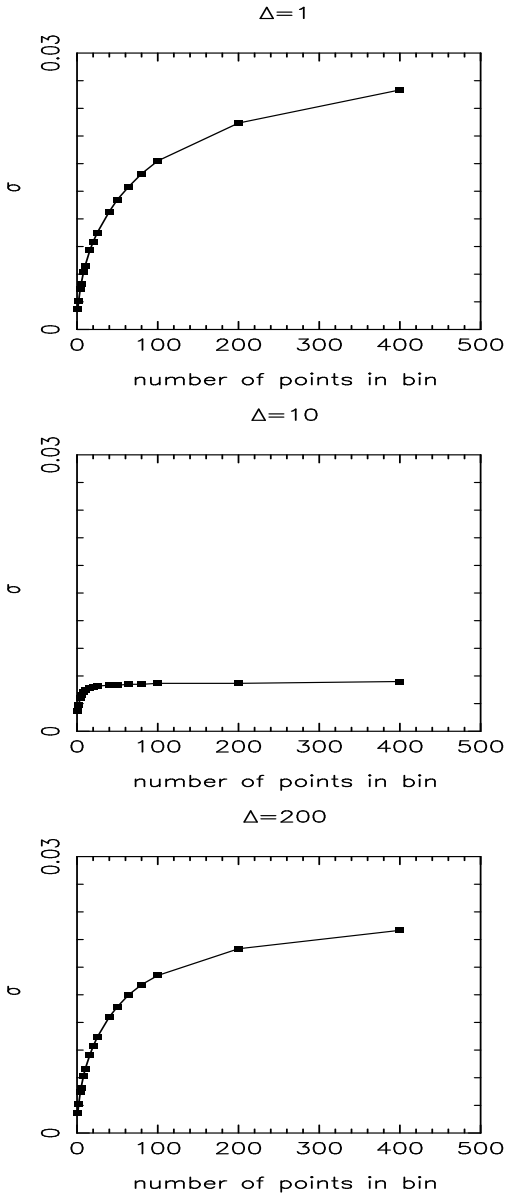

FIG. 3: Error bars for different bin sizes M. Data is that of Figures 1,2: step sizes $\Delta = 1, 10, 200$. (Acceptance rates=0.70, 0.35, 0.02).

Why does σ not increase indefinitely as M increases? You might expect it to, since the denominator $\sqrt{L-1}$ is getting smaller and smaller. The answer is that as more measurements are put in each bin, the different bins fluctuate less and less. The numerator which measures these fluctuations decreases in exact compensation to the denominator. (However, to reiterate, initially for M small when you put more measurements in the bins the new values are not independent and so the numerator does *not* decrease.)

• *So a second way to ensure the value for the error bar σ is correct is to consider different binnings of the data, and use the value obtained asymptotically as each bin contains a lot of data.*

How do we see this result is consistent with the correlation function analysis? There are two checks. The first is to see that the value for M at which σ starts to flatten out should be roughly the same as the value of τ for which $c(l)$ gets small. Second, one can compare the values of σ_1 at $M = 1$, where one assumes all the x are independent, with the asymptotic value σ_∞. The claim is that these should be related by $\sigma_\infty = \sqrt{\tau}\sigma_1$. You can see this is roughly true: For $\Delta = 1$ we get $\sigma_1 = 0.0022$ and $\sigma_\infty = 0.030$ from Figure 3. If you assume all the measurements are independent, you underestimate σ by more than an order of magnitude. Meanwhile, from figure 2, $\tau \approx 85$, and hence $\sqrt{\tau}$ similarly reflects this order of magnitude correction factor.

V. ACCEPTANCE RATE

The acceptance rate provides a rough guide to the choice of a good step size. If the acceptance rate is too much greater than 0.5, then one is likely in the limit of Figures (1–3)a where the correlation time is unnecessarily large due to small suggested moves. Likewise, if the acceptance rate is too much less than 0.50, then one is likely in the limit of Figures (1–3)c where the correlation time is unnecessarily large due to multiple rejections.

VI. MY "CHEAP" APPROACH

I recommended a "cheapo" approach to error bars, which was to bin the data from your run into 10 bins. ($M = N/10$.) This strategy assumes that you were doing a reasonably long run, so that $N/10 > \tau$. My thinking is that if this is violated, you are in more serious trouble than just getting the wrong error bars: you will have less than 10 independent measurements (and perhaps have not even properly equilibrated) so it is likely your expectation values themselves are wrong. For the particular example we are doing, the reported values from my simplistic approach for $\langle x^2 \rangle$ and σ were $1.973 \pm 0.051, 1.993 \pm 0.010, 1.995 \pm 0.039$ for $\Delta = 1, 10, 200$ respectively. These error bars should be the same as the asymptotic values of σ in Figure 3.

VII. FINAL COMMENTS

Does it matter which measurement you look at? I have looked entirely at x_i in doing this analysis. Would it matter if I had examined x_i^2 or some other measurement? For this simple problem I don't think so. In more complicated simulations it may be important to calculate correlation times separately for measurements of "local"

quantities (observables for degrees of freedom that are close together spatially) and "global" quantities (observables for degrees of freedom that are far apart spatially). The spatial separation of the different degrees of freedom in an observable can affect the autocorrelation time. In particular, observables containing quantities which are widely spaced generally have longer correlation times.

More generally, the size of the system you are simulating (and hence the spatial separations of the degreees of freedom) can greatly affect the correlation time, especially in the vicinity of a phase transition. Just as physical quantities like the specific heat, susceptibility etc can diverge at a critical point, the correlation time can diverge too as the system size increases near a critical point. This of course makes Monte Carlo very difficult. There is a big literature on solving this problem.