Statistical errors in histogram reweighting

Alan M. Ferrenberg,[1,2] D. P. Landau,[2] and Robert H. Swendsen[3]
[1]*University Computing and Networking Services, The University of Georgia, Athens, Georgia 30602*
[2]*Center for Simulational Physics, The University of Georgia, Athens, Georgia 30602*
[3]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*
(Received 14 July 1994)

Histogram reweighting methods now play an important role in Monte Carlo simulations, particularly in the study of critical phenomena. Despite this widespread use, a quantitative study of the statistical and systematic errors present when Monte Carlo data are reweighted has been lacking. In this paper, we present a detailed analysis of the statistical errors in histogram reweighting. The formalism is tested with simulations of the $d=2$ Ising model at infinite temperature and at the critical temperature. The error determined with this formalism agrees well with that calculated in the standard way of analyzing independent histograms. The implications of these results for high-resolution Monte Carlo studies are discussed.

PACS number(s): 02.70.Lq, 64.60.Cn, 05.70.Fh

I. INTRODUCTION

The resolution of Monte Carlo (MC) computer simulations has steadily increased as computers have become more powerful and simulation techniques more refined [1]. At the same time, our understanding of the potential sources of error in a simulation has also grown. The computation of statistical errors in a MC run, and the way in which these errors depend on the length of the run, as well as the correlation between successive configurations, has been studied in detail, and there is now a well-developed formalism for determining errors in thermodynamic quantities calculated directly from the simulations [2–4].

The development of histogram (reweighting) techniques [5,6] has allowed us to push the analysis of Monte Carlo data much farther than was previously thought possible [7]. Since the method is generally applicable to simulations of a wide variety of systems, the question of error determination in a histogram analysis is a problem of far ranging significance and interest. Results obtained by reweighting are subject to *systematic* errors due to the limited range of energies observed in a simulation, as well as an amplification of the normal *statistical* errors present due to the finite number of measurements made. In previous work, histograms from several different simulations (or equivalently different portions of a single, long simulation) were analyzed independently, and the variation of the results from the different analyses was used to estimate the error. While this is a practical and relatively efficient way to estimate errors (and in fact is the most commonly used method of error analysis in MC even when histogram reweighting is not used) it lacks the well-developed formalism which exists for standard MC simulations.

In this paper, we examine the question of the determination of the true statistical error of reweighed histogram data from a single run or histogram. In the following section we describe the theoretical formalism which can be used to determine the statistical error, and in Secs. III and IV we present results for several simple cases which can be used to test this approach. We close with some general remarks in the final section.

II. THEORY

Before examining the effect of statistical errors in histogram reweighting, it is instructive to review our understanding of errors in MC simulations. Because of the finite number of measurements, any quantity measured in a simulation will suffer from statistical and systematic errors [4]. This is further complicated by the fact that the measurements are not, in general, independent. The first careful study of statistical errors in MC simulations was performed by Müller-Krumbhaar and Binder [2] more than 20 years ago. They considered the statistical error in the average value of some quantity f measured in a simulation. If f_i is the value of f at the ith step of the simulation, the average value of f, $\langle f \rangle$, computed from a simulation consisting of N measurements (after discarding a sufficient number of measurements for equilibration), is

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f_i \ . \tag{1}$$

To calculate the statistical error in $\langle f \rangle$, Müller-Krumbhaar and Binder started with the expression for the variance $\mathcal{V}$ of a sum of N correlated random variables [8]

$$\mathcal{V} = N(\langle f^2 \rangle - \langle f \rangle^2) + 2 \sum_{i,j>i} \mathrm{cov}(f_i, f_j) \tag{2}$$

and related the covariance term to a sum of time-displaced averages. Their final expression is traditionally written as

$$(\delta f)^2 = (\langle f^2 \rangle - \langle f \rangle^2) \frac{2\tau_{f,f}+1}{N} \ , \tag{3}$$

where $\tau_{f,f}$ is the integrated correlation time [9] for the quantity f

$$\tau_{f,f}=\sum_{t=1}^{N}\left[1-\frac{t}{N}\right]\phi_{f,f}(t)$$

and $\phi_{f,f}(t)$ is the time-displaced autocorrelation function

$$\phi_{f,f}(t)=\frac{\langle f(0)f(t)\rangle-\langle f\rangle^2}{\langle f^2\rangle-\langle f\rangle^2}\ .$$

Our description of statistical errors in histogram reweighting will follow the Müller-Krumbhaar–Binder formalism rather closely. To see how this is possible, we first point out that MC data can be reweighted without using histograms. Consider a MC simulation performed at $\beta=\beta_0=1/kT_0$. The average value of some quantity $f(E)$ calculated using the single histogram method is

$$\langle f\rangle_\beta=\frac{\sum_E f(E)H(E)e^{-\Delta\beta E}}{\sum_E H(E)e^{-\Delta\beta E}}\ , \tag{4}$$

where $\Delta\beta=\beta-\beta_0$ and E is the *total* energy of a configuration. The histogram $H(E)$ is constructed from the time sequence of energies generated during the simulation

$$H(E)=\sum_{i=1}^{N}\delta(E,E_i)\ , \tag{5}$$

where δ is the Kronecker delta function and the sum runs over the N measurements made during the simulation. By inserting the definition of $H(E)$ (5), into (4), and performing the sum over E first, we get the equation for "reweighting on the fly," or reweighting without histograms:

$$\langle f\rangle_\beta=\frac{\sum_{i=1}^{N} f(E_i)e^{-\Delta\beta E_i}}{\sum_{i=1}^{N} e^{-\Delta\beta E_i}}=\frac{\mathcal{N}(\Delta\beta)}{\mathcal{D}(\Delta\beta)}\ . \tag{6}$$

When $\Delta\beta=0$, this reduces to the standard expression for the average of a quantity (1). The "reweighting on the fly" approach is useful for analyzing data requiring a multidimensional histogram, or for continuous systems to avoid the need to bin the data.

To simplify the formalism, we define a "curly bracket" notation for averages that include the reweighting factor $\exp[-\Delta\beta E_i]$. Each term inside the curly brackets carries along a reweighting factor. Examples of this notation are

$$\{f\}=\frac{1}{N}\sum_{i=1}^{N} f(E_i)e^{-\Delta\beta E_i}=\frac{\mathcal{N}(\Delta\beta)}{N}\ ,$$

$$\{1\}=\frac{1}{N}\sum_{i=1}^{N} e^{-\Delta\beta E_i}=\frac{\mathcal{D}(\Delta\beta)}{N}\ ,$$

$$\{f,1\}=\frac{1}{N}\sum_{i=1}^{N} f(E_i)e^{-2\Delta\beta E_i}\ .$$

Note that these averages can also be calculated using the corresponding histograms. The single-histogram equation itself (4) expressed in this notation is

$$\langle f\rangle_\beta=\frac{\{f\}}{\{1\}}\ .$$

The analysis of errors is complicated by the fact that once reweighting has been performed, both the numerator $\mathcal{N}(\Delta\beta)$ and denominator $\mathcal{D}(\Delta\beta)$ in (6) will suffer from statistical error [in (1), the denominator is simply the number of measurements, N, which has no error]. We represent the square of these errors by $\delta\mathcal{N}^2(\Delta\beta)$ and $\delta\mathcal{D}^2(\Delta\beta)$, respectively. In addition, we expect that the error in the numerator is correlated with that of the denominator because both are calculated from the same set of measurements. It is important to note that this correlation is present *even if* there is no correlation between measurements during the simulation.

If $\mathcal{N}(\Delta\beta)$ and $\mathcal{D}(\Delta\beta)$ were independent, the square of the statistical error in $\langle f\rangle_\beta$ would be given by

$$(\delta f)^2=\left[\frac{\partial\langle f\rangle_\beta}{\partial\mathcal{N}(\Delta\beta)}\right]^2\delta\mathcal{N}^2(\Delta\beta)+\left[\frac{\partial\langle f\rangle_\beta}{\partial\mathcal{D}(\Delta\beta)}\right]^2\delta\mathcal{D}^2(\Delta\beta)$$

or

$$(\delta f)^2=\frac{1}{\mathcal{D}(\Delta\beta)^2}\delta\mathcal{N}^2(\Delta\beta)+\frac{\mathcal{N}(\Delta\beta)^2}{\mathcal{D}(\Delta\beta)^4}\delta\mathcal{D}^2(\Delta\beta)\ , \tag{7}$$

which is the standard expression for the propagation of error in a function of two independent variables [10]. However, because $\mathcal{N}(\Delta\beta)$ and $\mathcal{D}(\Delta\beta)$ are not independent, (7) is not correct, and in fact overestimates the true error. To properly take this correlation into account, we must include the covariance $\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)$. The correct expression for the square of the statistical error in $\langle f\rangle_\beta$ is then given by

$$(\delta f)^2=\left[\frac{\partial\langle f\rangle_\beta}{\partial\mathcal{N}(\Delta\beta)}\right]^2\delta\mathcal{N}^2(\Delta\beta)+\left[\frac{\partial\langle f\rangle_\beta}{\partial\mathcal{D}(\Delta\beta)}\right]^2\delta\mathcal{D}^2(\Delta\beta)$$
$$-2\left[\frac{\partial^2\langle f\rangle_\beta}{\partial\mathcal{N}(\Delta\beta)\partial\mathcal{D}(\Delta\beta)}\right]\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)$$

or

$$(\delta f)^2=\frac{1}{\mathcal{D}(\Delta\beta)^2}\delta\mathcal{N}^2(\Delta\beta)+\frac{\mathcal{N}(\Delta\beta)^2}{\mathcal{D}(\Delta\beta)^4}\delta\mathcal{D}^2(\Delta\beta)$$
$$-\frac{2\mathcal{N}(\Delta\beta)}{\mathcal{D}(\Delta\beta)^3}\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)\ .$$

The square of the *relative error* in $\langle f\rangle_\beta$ takes on a particularly simple form:

$$\frac{(\delta f)^2}{\langle f\rangle_\beta^2}=\frac{\delta\mathcal{N}^2(\Delta\beta)}{\mathcal{N}(\Delta\beta)^2}+\frac{\delta\mathcal{D}^2(\Delta\beta)}{\mathcal{D}(\Delta\beta)^2}-\frac{2\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)}{\mathcal{N}(\Delta\beta)\mathcal{D}(\Delta\beta)}\ . \tag{8}$$

To facilitate this derivation, let us consider the covariance $\delta R\,\delta Q$ of two arbitrary functions R and Q which have the same form as $\mathcal{N}(\Delta\beta)$ and $\mathcal{D}(\Delta\beta)$:

$$R=\sum_{i=1}^{N} r(E_i)e^{-\Delta\beta E_i}=N\{r\}\ ,$$

$$Q=\sum_{i=1}^{N} q(E_i)e^{-\Delta\beta E_i}=N\{q\}\ .$$

From this, we can then easily calculate $\delta\mathcal{N}^2(\Delta\beta)$, $\delta\mathcal{D}^2(\Delta\beta)$, and $\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)$ by replacing the functions r and q with f and 1 appropriately. By generalizing the analysis of Müller-Krumbhaar and Binder, we can define the covariance $\delta R\,\delta Q$ as

$$\delta R\,\delta Q=N(\{r,q\}-\{r\}\{q\})+2\sum_{i,j>i}\mathrm{cov}[r(E_i)e^{-\Delta\beta E_i},q(E_j)e^{-\Delta\beta E_j}]\ . \tag{9}$$

The double sum over i and j can be replaced by a single sum of time-displaced averages

$$2N\sum_{t=1}^{N-1}\left[1-\frac{t}{N}\right](\{r(0),q(t)\}-\{r\}\{q\})\ , \tag{10}$$

where t is the time displacement index. The covariance (9) is thus expressed as

$$\delta R\,\delta Q=N(\{r,q\}-\{r\}\{q\})\times\left[1+2\sum_{t=1}^{N-1}\left[1-\frac{t}{N}\right]\frac{\{r(0),q(t)\}-\{r\}\{q\}}{\{r,q\}-\{r\}\{q\}}\right]\ . \tag{11}$$

To complete the generalization of the Müller-Krumbhaar-Binder formalism, we define a reweighted time-displaced cross-correlation function

$$\phi_{\{r\},\{q\}}(t)=\frac{\{r(0),q(t)\}-\{r\}\{q\}}{\{r,q\}-\{r\}\{q\}}$$

and a reweighted correlation time

$$\tau_{\{r\},\{q\}}=\sum_{t=1}^{N}\left[1-\frac{t}{N}\right]\phi_{\{r\},\{q\}}(t) \tag{12}$$

to finally obtain

$$\delta R\,\delta Q=N(\{r,q\}-\{r\}\{q\})(1+2\tau_{\{r\},\{q\}})\ . \tag{13}$$

With the proper substitutions for r and q in (13) we can now evaluate $\delta\mathcal{N}^2(\Delta\beta)$, $\delta\mathcal{D}^2(\Delta\beta)$, and $\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)$:

$$\delta\mathcal{N}^2(\Delta\beta)=N(\{f,f\}-\{f\}^2)(1+2\tau_{\{f\},\{f\}})\ ,$$

$$\delta\mathcal{D}^2(\Delta\beta)=N(\{1,1\})-\{1\}^2)(1+2\tau_{\{1\},\{1\}})\ ,$$

$$\delta\mathcal{N}\delta\mathcal{D}(\Delta\beta)=N(\{f,1\}-\{f\}\{1\})(1+2\tau_{\{f\},\{1\}})$$

and the relative error in f

$$\frac{(\delta f)^2}{\langle f\rangle_\beta^2}=\frac{1}{N}\left[\left[\frac{\{f,f\}}{\{f\}^2}-1\right](1+2\tau_{\{f\},\{f\}})+\left[\frac{\{1,1\}}{\{1\}^2}-1\right](1+2\tau_{\{1\},\{1\}})-2\left[\frac{\{f,1\}}{\{f\}\{1\}}-1\right](1+2\tau_{\{f\},\{1\}})\right]\ . \tag{14}$$

The error in a thermodynamic quantity like the energy can be obtained simply by replacing f with E in (14). To compute the error in a response function, for example, the specific heat, we need to find the appropriate function $f(E)$ whose average value gives us the desired quantity. For the specific heat, the function is

$$f(E)=(E-\langle E\rangle_\beta)^2$$

so that the specific heat C is given by

$$C=\frac{\beta^2}{V}\langle f(E)\rangle_\beta\ .$$

The quantities of interest are then

$$\{f,f\}=\{E^4,1\}-4\langle E\rangle_\beta\{E^3,1\}+6\langle E\rangle_\beta^2\{E^2,1\}-4\langle E\rangle_\beta^3\{E,1\}+\langle E\rangle_\beta^4\{1,1\}\ ,$$

$$\{f,1\}=\{E^2,1\}-2\langle E\rangle_\beta\{E,1\}+\langle E\rangle_\beta^2\{1,1\}\ ,$$

$$\{f\}=\{E^2\}-\langle E\rangle_\beta^2\{1\}\ .$$

The time-displaced correlation functions are quite complex. For example, to calculate

$$\phi_{\{f\},\{f\}}(t)=\frac{\{f(0),f(t)\}-\{f\}^2}{\{f,f\}-\{f\}^2}$$

for the specific heat, we need $\{f(0),f(t)\}-\{f\}^2$ which is given by

$$\phi_{\{E^2\},\{E^2\}}(t)\mathrm{cov}_{\{\}}(E^2,E^2)-4\langle E\rangle_\beta\phi_{\{E^2\},\{E\}}(t)\mathrm{cov}_{\{\}}(E^2,E)+2\langle E\rangle_\beta^2\phi_{\{E\},\{E\}}(t)\mathrm{cov}_{\{\}}(E,E)$$
$$+4\langle E\rangle_\beta^2\phi_{\{E^2\},\{1\}}(t)\mathrm{cov}_{\{\}}(E^2,1)-4\langle E\rangle_\beta^3\phi_{\{E\},\{1\}}(t)\mathrm{cov}_{\{\}}(E,1)+\langle E\rangle_\beta^4\phi_{\{1\},\{1\}}(t)\mathrm{cov}_{\{\}}(1,1)$$

and $\{f,f\}-\{f\}^2$ which is given by

$$\mathrm{cov}_{\{\}}(E^2,E^2)-4\langle E\rangle_\beta\mathrm{cov}_{\{\}}(E^2,E)+2\langle E\rangle_\beta^2\mathrm{cov}_{\{\}}(E,E)+4\langle E\rangle_\beta^2\mathrm{cov}_{\{\}}(E^2,1)-4\langle E\rangle_\beta^3\mathrm{cov}_{\{\}}(E,1)+\langle E\rangle_\beta^4\mathrm{cov}_{\{\}}(1,1)\ ,$$

where

$$\mathrm{cov}_{\{\}}(X,Y)=\{X,Y\}-\{X\}\{Y\} .$$

A different approach to calculating the error in C is to make two independent passes through the data, calculating $\langle E\rangle_\beta$ in the first pass, then directly evaluating $(E-\langle E\rangle_\beta)^2$ in the second. We found that this second approach is easier to implement, and is more stable numerically.

The expressions for the error have two different kinds of terms: some that depend on the simulation algorithm used, containing the reweighted correlation times, and others that represent equilibrium averages and are therefore independent of the simulation algorithm. However, unlike the nonreweighted case, we cannot simply factor out the correlation time dependence; this will lead to nontrivial differences in how the error increases with $\Delta\beta$ when we change from one simulation algorithm to another! Examples making use of the formalism developed here are given in the next two sections.

III. GAUSSIAN TEST CASE

In this section, we apply the formalism for the statistical error to the case of a Gaussian energy distribution with independent measurements. While it is clear that no real physical system is described by such a distribution, we nevertheless feel it is instructive to carry out the calculation because it allows us to derive a simple, closed-form expression for the statistical error. In addition, we expect that the behavior of this "toy" system will be qualitatively the same as that of a real system.

With independent measurements, all correlation times are zero, and the expression for the relative error (14) is simplified to be

$$\frac{\delta f^2}{\langle f\rangle_\beta^2}=\frac{1}{N}\left[\frac{\{f,f\}}{\{f\}^2}+\frac{\{1,1\}}{\{1\}^2}-2\frac{\{f,1\}}{\{f\}\{1\}}\right] . \tag{15}$$

For this example, we will use a continuous probability distribution which is symmetric about $E=0$ and has a standard deviation σ:

$$P(E)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{E^2}{2\sigma^2}\right] .$$

The quantities needed for the error analysis are then easily computed by integrating $P(E)$ with the appropriate function. For this example, we will calculate the statistical error in the energy and specific heat so that the quantities we need are

$$\{1\}=\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right] ,$$

$$\{E\}=-\Delta\beta\sigma^2\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right] ,$$

$$\{E^2\}=(\sigma^2+\Delta\beta^2\sigma^4)\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right] ,$$

$$\{E^3\}=-(3\Delta\beta\sigma^4+\Delta\beta^3\sigma^6)\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right] ,$$

$$\{E^4\}=(3\sigma^4+6\Delta\beta^2\sigma^6+\Delta\beta^4\sigma^8)\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right] ,$$

$$\{1,1\}=\exp[2\Delta\beta^2\sigma^2] ,$$

$$\{E,1\}=-2\Delta\beta\sigma^2\exp[2\Delta\beta^2\sigma^2] ,$$

$$\{E^2,1\}=(\sigma^2+4\Delta\beta^2\sigma^4)\exp[2\Delta\beta^2\sigma^2] ,$$

$$\{E^3,1\}=-(6\Delta\beta\sigma^4+8\Delta\beta^3\sigma^6)\exp[2\Delta\beta^2\sigma^2] ,$$

$$\{E^4,1\}=(3\sigma^4+24\Delta\beta^2\sigma^6+16\Delta\beta^4\sigma^8)\exp[2\Delta\beta^2\sigma^2] .$$

By inserting these into (15) we find that the expression for the square of the relative error in $\langle E\rangle_\beta$ is

$$\frac{(\delta E)^2}{\langle E\rangle_\beta^2}=\frac{\exp[\Delta\beta^2\sigma^2]}{N}\left[1+\frac{1}{\Delta\beta^2\sigma^2}\right] .$$

Because the average value of E is zero at $\Delta\beta=0$, the relative error is not well behaved. Let us therefore consider the absolute statistical error, δE:

$$\delta E=\frac{\sigma}{\sqrt{N}}\exp\left[\frac{\Delta\beta^2\sigma^2}{2}\right]\sqrt{1+\Delta\beta^2\sigma^2} . \tag{16}$$

For $\Delta\beta=0$, the error reduces to the expected result $\sigma/\sqrt{N}$. For small values of $\Delta\beta$, the error increases moderately, but as $\Delta\beta$ gets larger, the error in E begins to increase dramatically due to the exponential term $\exp[\Delta\beta^2\sigma^2/2]$ and continues to rise without limit. The error in the specific heat has nearly the same form as the error in the energy:

$$\frac{\delta C}{C}=\frac{\exp[\Delta\beta^2\sigma^2/2]}{\sqrt{N}}\sqrt{2+4\Delta\beta^2\sigma^2+\Delta\beta^4\sigma^4} . \tag{17}$$

While the explicit form will be different for real systems, the qualitative behavior of the error should remain: For a region around $\Delta\beta=0$, the errors will increase only moderately, then will begin to rise rapidly as $\Delta\beta$ increases.

IV. $d=2$ ISING MODEL

To further test the formalism developed in Sec. II, we need to study a more physically interesting case. The $d=2$ Ising model [11,12] is especially attractive because we can compare the calculated errors with the true deviation from the exactly known answer for a finite system [13]. In addition, we can study the effects of correlated measurements by using both the standard Metropolis algorithm [14] and the Wolff algorithm [15] in the simulations. We will consider the Ising model with the Hamiltonian

$$\mathcal{H}=-J\sum_{\langle i,j\rangle}\sigma_i\sigma_j=-JE ,$$

where J is the ferromagnetic coupling constant (which we set equal to 1), σ takes on the values ±1, and the sum

runs over all nearest-neighbor pairs. E is the dimensionless total energy.

The simulations were carried out on square lattices with linear dimension $L=32$ and periodic boundary conditions in both lattice directions. Simulations were performed at infinite temperature ($\beta=0$) and at the critical temperature T_c [12]. The results are described in the next two sections.

A. Infinite temperature Ising model

For ($\beta=0$), we expect the results to be similar to the Gaussian test case since all correlation times are zero, and the distribution is approximately Gaussian. (In addition, the standard deviation σ used in the Gaussian test case was chosen to be that of a 32×32 Ising model at $\beta=0$.) We performed 25 simulations with 10^6 MCS (complete lattice updates) per simulation. Because $\beta=0$, this corresponds to a simple-sampling rather than an importance-sampling simulation. The measured absolute error in the average energy $\langle E\rangle$ and relative error in the specific heat C were determined by reweighting the histograms independently, then considering the distribution of the calculated values of $\langle E\rangle$ and C at each value of β. The uncertainty in the error was estimated by breaking the data up into five bins of five simulations each. We believe that our values for the relative error as a function of β will be well defined, but our estimate of the uncertainty in the error could be off by a factor of 2. To calculate the theoretical error, we evaluated (15) as a function of β for each of the 25 simulations.

In Fig. 1, the measured statistical error in the energy is plotted along with the prediction from (15). The result for the Gaussian test case using the standard deviation of the energy measured in the simulation for σ in (16) is also shown. The agreement is very good from $\beta=0$ to around $\beta=0.1$, a range which encompasses more than two orders of magnitude in the error. By the time this value of β is reached, both the theoretical and measured values for the error suffer from *systematic errors* due to the reweighting procedure and are unreliable—the true error should continue to rise, as it does in the Gaussian test case. Figure 2 compares the theoretical and measured relative error in the specific heat for the same data as in Fig. 1.

Since successive configurations generated from a MC simulation at infinite temperature are independent of each other, there might be a temptation to believe that the reweighting of quantities to the critical region would not suffer from increased statistical errors due to critical slowing down. This temptation would even appear to be supported by a naive calculation of a correlation time which is zero for simple sampling, even in the presence of reweighting. However, the generation of useful configurations is, in fact, subject to an extremely strong, effective "slowing down," which renders such simulations of very limited value. Indeed, it was the introduction of importance sampling by Metropolis *et al.* [14] that improved efficiency to make the Monte Carlo method practical.

The "slowing down" due to infinite temperature simulations can be seen especially clearly in our example of the Gaussian distribution. The probability of generating configurations with values of the energy near $\langle E\rangle=-\Delta\beta\sigma^2$ decreases as $\exp[-\langle\Delta\beta\rangle^2\sigma^2/2]$. Therefore, when the energy of interest is far from the average energy of the simulation in units of the width of the histogram distribution, the number of useful configurations becomes an extremely small fraction of the total number of generated configurations. The ratio of the width of the histogram distribution from an infinite temperature simulation to the energy separation goes to zero as the square root of the volume of the system, which is clearly reflected in our expression for the error. This "slowing down" is far worse than any critical slowing down due to any standard MC methods.

This error also occurs in a somewhat more subtle apparent absence of slowing down in a recent paper by Hu [16]. Hu applied the Kastelyn-Fortuin [17] mapping between Potts models and percolation models to use simulations of a percolation model to extrapolate to the properties of the Ising model. The method is, of course, valid, but the claim of avoiding slowing down is incorrect for

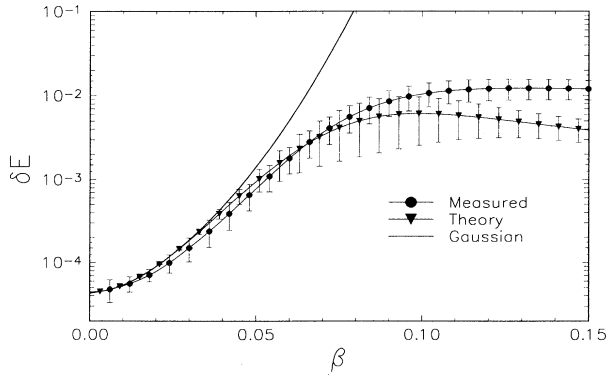

FIG. 1. Absolute error in E as a function of β for the $d=2$ Ising model simulated at infinite temperature ($\beta=0$). The measured error is compared to the theoretical error from (15) as well as the Gaussian test model (16).

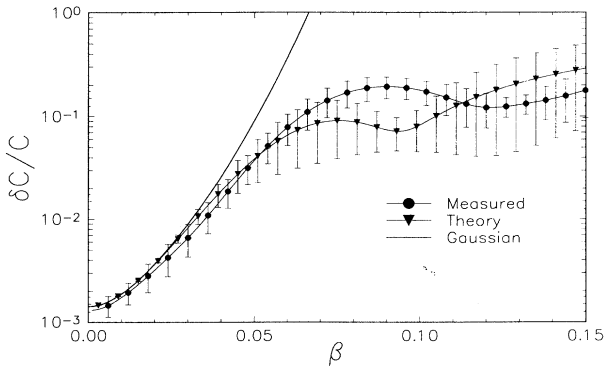

FIG. 2. Relative error in C as a function of β for the $d=2$ Ising model simulated at infinite temperature ($\beta=0$). The measured error is compared to the theoretical error from (15) as well as the Gaussian test model (17).

the reasons given above. The extrapolation involves a weighting factor of s^{N_c}, where s is the number of Potts states and N_c is the number of clusters. For Hu's application, $s=2$, and N_c is proportional to the volume of the system. Therefore, the fraction of useful configurations decreases exponentially, and the errors increase exponentially with the volume of the system. The apparent avoidance of critical slowing down occurs at the price of a much more severe inefficiency.

B. Ising model at T_c

The Ising model at T_c is a more interesting case than the two previously discussed. Measurements made during a simulation at $T=T_c$ may be strongly correlated, so the full formalism (14) must be applied to describe the statistical errors. More importantly, the critical region is where the histogram method is of most practical use, so understanding the errors near T_c is of more general interest.

To study the Ising model at T_c, three different groups of simulations were performed:

(i) 10 simulations of length 10^6 MCS using the Metropolis algorithm, and keeping the time series for correlation time analysis;

(ii) 60 simulations of length 10^6 MCS using the Metropolis algorithm, keeping only histograms;

(iii) 5 simulations of 10^6 updates performed using the Wolff algorithm, and keeping the time series for correlation time analysis.

For the Metropolis algorithm study, we calculated the measured error using the group of 60 simulations and the theoretical error using the group of 10 simulations. The reason for this is twofold. First, by computing the measured and theoretical errors from different simulations, we eliminate the possibility of a false agreement between "experiment" and theory due to some subtle correlation. Second, the computation of the theoretical error is quite CPU intensive. We estimate than an additional 13 days of CPU time (on an IBM RISC/6000 model 550 workstation) would have been required to perform the complete correlation time analysis for the extra 60 simulations. The uncertainty in the measured error was estimated by breaking the data up into 6 bins of 10 simulations each. Because this is a difficult process, we expect that our estimate for the error in the error could be off by 50% or more, although the estimate for the error itself is quite reliable. For the Wolff algorithm study, only the theoretical errors were determined.

The computation of the theoretical error is quite challenging due to strong cancellation effects between the three terms in (14) as well as numerical instabilities in the determination of the reweighted correlation time (12). To overcome these problems, it was necessary to rewrite (14) in a more stable form. The numerically stable expression for the first term is

$$\frac{\{f,f\}}{\{f\}^2}+2\sum_{t=1}^{N}\left[1-\frac{t}{N}\right]\frac{\{f(0),f(t)\}}{\{f\}^2}$$

with similar modifications to the second and third terms.

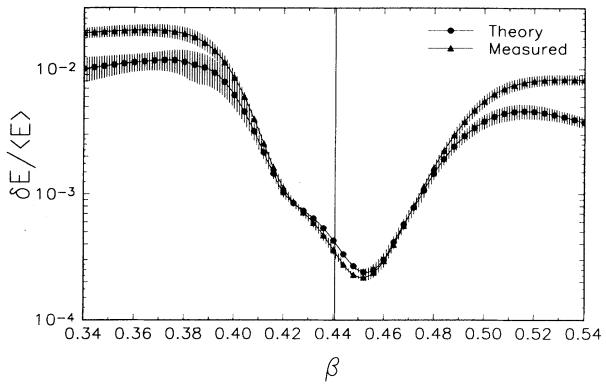

FIG. 3. Comparison of the theoretical and measured relative error in E as a function of β for the $d=2$ Ising model simulated at T_c (marked by the vertical line). The simulations were performed with the Metropolis algorithm.

Figures 3 and 4 compare the theoretical and measured error in the energy and specific heat (respectively) for the Metropolis algorithm simulations. The agreement between theory and experiment is very good for a range of β values around β_0. By the time the two error estimates disagree, the error has increased by more than an order of magnitude, and the results obtained are no longer reliable due to systematic errors from the reweighting. Further evidence of this is given in Fig. 5, which shows the ratio of ΔE, the deviation of the average energy from the exactly known value, to the measured error δE. The measured energy agrees with the exact energy, within the calculated error, for β values in the range 0.38–0.51 which roughly corresponds to the range in which the theoretical and measured errors agree.

In Fig. 3, we can also see the curious result that the minimum error does not occur at the β value where the simulation was performed. This results from the fact that the measurements are correlated, and that the correlation

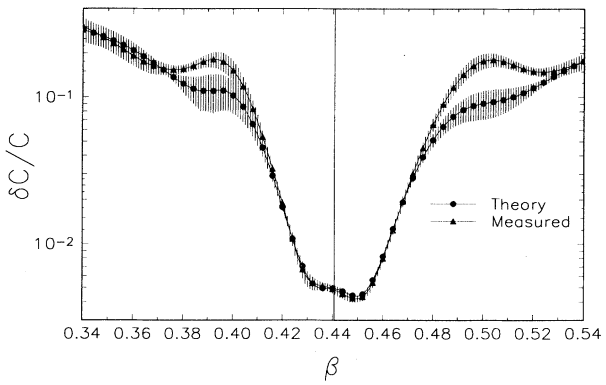

FIG. 4. Comparison of the theoretical and measured relative error in C as a function of β for the $d=2$ Ising model simulated at T_c. The vertical line indicates the simulated temperature. Results obtained using the Metropolis algorithm are shown.

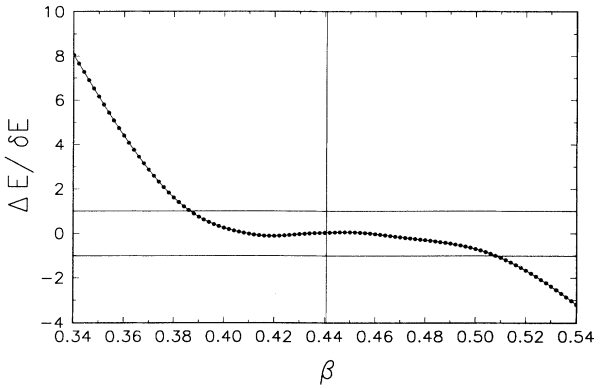

FIG. 5. Ratio of the deviation from the correct answer to the statistical error for the $d=2$ Ising model simulated at T_c (marked by the vertical line) using the Metropolis algorithm. The horizontal lines represent ± 1 standard deviation.

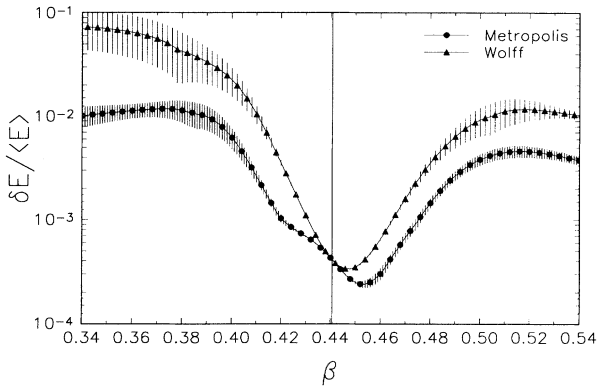

FIG. 7. Comparison of the relative error in E determined by Metropolis and Wolff simulations for the $d=2$ Ising model. To simplify the comparison, the error from the Wolff algorithm was rescaled to match that of the Metropolis algorithm at the simulated temperature (marked by the vertical line).

time is larger for β values below β_c. If we assume that our equations for the error are correct, we can use the Metropolis histograms to estimate what the error would be if all the measurements were, in fact, uncorrelated. That is, we can evaluate (15) using the Metropolis histograms. (In an actual simulation, independent samples could be obtained by discarding a number of MCS equal to the correlation time between making measurements.) The comparison of the measured errors with those expected for no correlations is shown in Fig. 6. To facilitate the comparison, the number of measurements for the uncorrelated data was adjusted so that the errors were the same at the simulated value of β. The minimum error for uncorrelated measurements is indeed almost exactly where the simulation was performed, and the increase in the error is fairly symmetric around β_0. An important consequence of this result is that when histogram reweighting is used in a MC study, *every* measurement made should go into constructing the histogram. Without reweighting, the statistical and systematic errors in quantities depend only on the number of independent measurements made. Figure 6 clearly shows that this is not the case for reweighting.

If the shift in the minimum is due to the correlation time, then we would expect results obtained using the Wolff algorithm to differ from the Metropolis results. In Fig. 7, we compare the theoretical error for the two algorithms, choosing the number of Wolff measurements so that the statistical error for the two algorithms at β_0 coincide. There is indeed a substantial difference in the temperature dependence of the error. The most important feature of this plot is the slower increase in the error for the Metropolis data, compared to the Wolff data, when they are reweighted. This phenomenon must be taken into account in studies of critical phenomena, where it is important to have reliable information over a range of

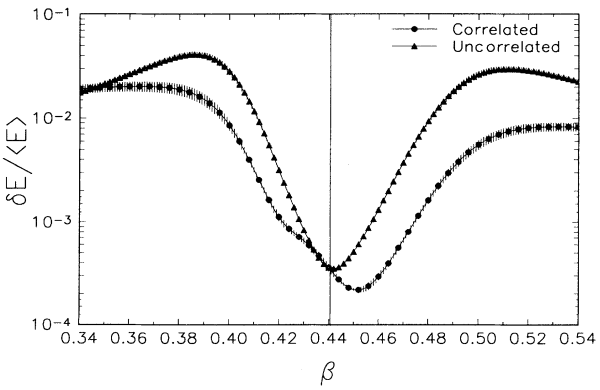

FIG. 6. Comparison of the relative error in the energy for a simulation with independent measurements to a Metropolis simulation with correlations. The vertical line indicates the location of the simulated temperature. Results for the $d=2$ Ising model are shown.

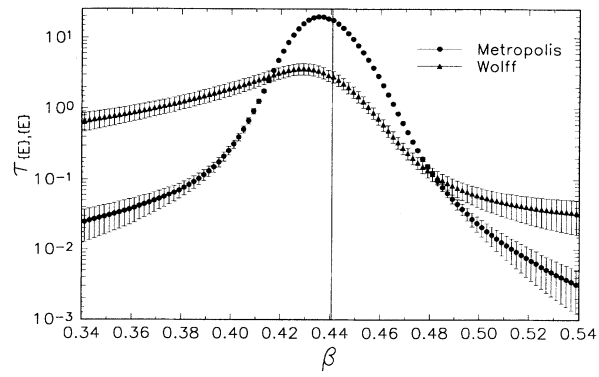

FIG. 8. Plot of the reweighted energy correlation time (12) for the Metropolis and Wolff algorithms. Results for the $d=2$ Ising model are shown. The simulations were performed at $T=T_c$, which is marked by the vertical line.

temperatures to locate peaks in thermodynamic functions.

The "reweighted" correlation time (12) for the energy is shown as a function of β in Fig. 8 for the Wolff and Metropolis algorithms. (The other correlation times exhibit similar behavior.) The temperature dependence of the reweighted correlation time is similar to that of the true correlation times: The correlation time for the Wolff algorithm remains fairly constant over the temperature range considered while τ for the Metropolis algorithm varies greatly as would be expected due to its larger value of the dynamic critical exponent z. It would be a most fascinating result if the real and reweighted times were the same, but this is, unfortunately, not the case.

V. DISCUSSION AND CONCLUSIONS

The problem of understanding statistical and systematic errors in histogram reweighting techniques is an important one. In this work, we have performed the first quantitative study of the statistical errors in reweighted data. The formalism which we have developed and tested now puts the entire question of error determination for reweighted data on a firm basis. The agreement between the measured and theoretical errors is quite good; indeed, it is better than we had anticipated. Both the measured and theoretical errors suffer from *systematic* errors due to the reweighting procedure: Rather than continuing to rise dramatically with $\Delta\beta$, they tend to either become constant, or at least increase at a much slower rate than they should. The point at which the error stops growing rapidly can therefore provide an estimate for the limit to the range of applicability of the reweighting method. The theoretical calculation of the error is both difficult and time consuming so it may not be justified for all studies. However, our goal was to show that there really is a theoretical foundation for the description of the errors, and our results justify the procedure of calculating errors from multiple independent runs.

This problem is also presently under study using a different approach by Janke [18]. Münger and Novotny [19] have investigated some systematic effects in reweighting by means of studying independent simulations, but they developed no formalism to describe their results. The widespread use of histogram reweighting has led to the development of several related analysis and simulation techniques. Multicanonical reweighting methods [20,21] have shown great promise for simulations of first-order transitions as well as of spin glasses. Janke and Sauer [22] have recently shown that the multicanonical method is subject to the same kinds of errors as histogram reweighting; in fact, the final result for the statistical error (14) is nearly identical for the two methods.

Of particular concern for high-resolution MC studies is the observation that the minimum statistical error in the temperature dependence of quantities calculated using histograms does not necessarily occur at the temperature where the simulation was performed. This is an effect of the correlation time, because the minimum error for uncorrelated data is found exactly where the simulation was performed. This effect must be taken into account when selecting a simulation algorithm for a particular model, and it will clearly be an important consideration in the development of simulation algorithms to be used to study critical phenomena with high resolution. In particular, cluster algorithms have been promoted in recent years as an efficient way of reducing correlation between configurations, and thus the error. However, since the critical temperature for an arbitrary model is not known in advance, it is likely that the simulation will not actually be performed at the critical point. As the temperature moves away from T_c, the correlation time decreases quite rapidly for the Metropolis method; simulations performed slightly away from the critical point become relatively more efficient than do cluster flipping studies. This suggests that hybrid Monte Carlo methods [23] which combine techniques to produce many configurations quite rapidly with only modest correlation between them might be particularly effective. Of course, different quantities have different correlation times and will thus have different errors in the histogram analysis just as in a simple Monte Carlo study. The variation of the error in the histogram analysis then allows us to estimate how computer time can be used most efficiently, i.e., either to improve the quality of the histogram for an initial temperature or by producing data at a shifted temperature.

ACKNOWLEDGMENTS

We wish to thank Wolfhard Janke and Bernd Berg for interesting discussions. This work was supported, in part, by NSF Grant No. DMR-9405018.

[1] For a review, see *The Monte Carlo Method in Condensed Matter Physics,* edited by K. Binder (Springer, Berlin, 1992).

[2] H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. **8**, 1 (1973).

[3] N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).

[4] A. M. Ferrenberg, D. P. Landau, and K. Binder, J. Stat. Phys. **63**, 867 (1991).

[5] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).

[6] R. H. Swendsen, J.-S. Wang, and A. M. Ferrenberg, in *The Monte Carlo Method in Condensed Matter Physics* (Ref. [1]).

[7] A. M. Ferrenberg and D. P. Landau, Phys. Rev. B **44**, 5081 (1991).

[8] An old but excellent reference is W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1957).

[9] The definition of the integrated correlation time used here differs from that used in some recent papers. See, for example, Ref. [3].

[10] See, for example, P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).

[11] E. Ising, Z. Phys. **31**, 2553 (1925).
[12] L. Onsager, Phys. Rev. **65**, 117 (1944).
[13] A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).
[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[15] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[16] C.-K. Hu, Phys. Rev. Lett. **69**, 2739 (1992).
[17] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. Suppl. **26s**, 11 (1969); C. M. Fortuin and P. W. Kasteleyn, Physica **57**, 536 (1972).
[18] W. Janke (unpublished).
[19] E. Münger and M. Novotny, Phys. Rev. B **43**, 5773 (1991).
[20] B. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[21] B. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).
[22] W. Janke and T. Sauer, Phys. Rev. E **49**, 3475 (1994).
[23] A. M. Ferrenberg and D. P. Landau (unpublished).