3
Data analysis

An important part of any Monte Carlo simulation is the statistical analysis of the data it generates. Central to this analysis is computing the average values of the relevant physical observables and estimating the statistical error of these averages. Computing the averages is easy. It mainly requires ensuring the simulation is sampling configurations from the equilibrium distribution before tallying the values of the observables. Estimating the error is more difficult. The increased difficulty is not in evaluating its mathematical definition, the so-called standard deviation, but rather it is in ensuring the measurements used in this formula are statistically independent. The correlations among the data points depend on the values of a model's parameters, the observable, and the sampling algorithm being used. We now discuss the statistical basis of common procedures for analyzing the results of a Monte Carlo simulation.

3.1 Equilibrating the sampling

We know from Section 2.4 that the Markov chain requires a certain number of steps before it begins to sample from its limiting distribution. This distribution is also called the *equilibrium distribution* even for cases where we are not simulating a problem in equilibrium statistical mechanics. Verifying that equilibrium has been reached is a generic task of all Markov chain Monte Carlo simulations. Once in equilibrium, we then want to use the chain samples to estimate the values of various physical quantities and their associated statistical errors.

There are physics-related issues that can make equilibration particularly difficult, many of which affect simulations in the vicinity of phase transitions. Although we might know what phases to expect, we generally do not know the precise location of phase boundaries. Identifying the phases and determining their locations in parameter space are common tasks for which Monte Carlo simulations are employed.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

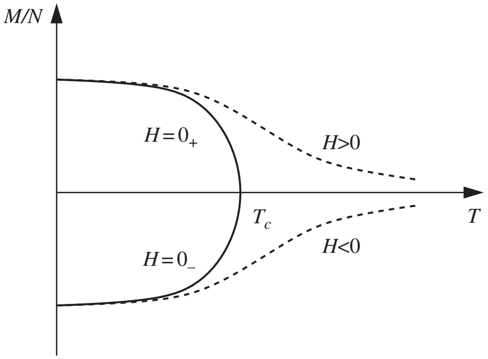

Figure 3.1 Sketch of the magnetization of the two-dimensional Ising model as a function of T for zero field (solid lines) and nonzero field (dashed lines). T_c is the critical temperature of the zero field case. The magnetization per site (M/N) is the order parameter.

The two-dimensional Ising model $E = -J\sum_{\langle ij\rangle} s_i s_j + H\sum_i s_i$, with $\langle ij\rangle$ denoting nearest neighbor sites, is a convenient example for illustrating the difficulty in equilibrating near a phase transition and the other main points of this chapter.

The magnetization of the two-dimensional Ising model as a function of temperature T is sketched in Fig. 3.1. In zero field, as the temperature is lowered, a second-order (continuous) phase transition occurs at the critical temperature $T_c = 2J/\log(1+\sqrt{2}) \approx 2.269J$, where the system transitions from a disordered to an ordered phase. In a nonzero field, at fixed $T < T_c$, a first-order transition occurs as we switch the sign of the field.

In the vicinity of either type of transition, Monte Carlo simulations, especially those based on local update Metropolis or heat-bath algorithms, experience difficulties in equilibrating. These algorithms move through phase space with small changes in configurations associated with small changes in energy. A first-order transition is marked by an energy barrier created by the surface tension between the two phases that is hard to surmount in a single or a chained couple of moves. The problem here is getting stuck in one part of phase space for a long time. A second-order phase transition is marked by critical slowing down. This phenomenon, which is related to a diverging correlation length, affects a Monte Carlo simulation if the updating procedure changes the configuration only locally. The problem here is knowing when the system has relaxed.

Before we can begin a proper analysis of the data we must first determine the number of Monte Carlo steps necessary to equilibrate. This number is generally called the *equilibration time*. Typically, we measure this time by the number of *sweeps*. One sweep is the result of a procedure that attempts at least one local Monte Carlo update of all the random variables. In the case of the Ising model,

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

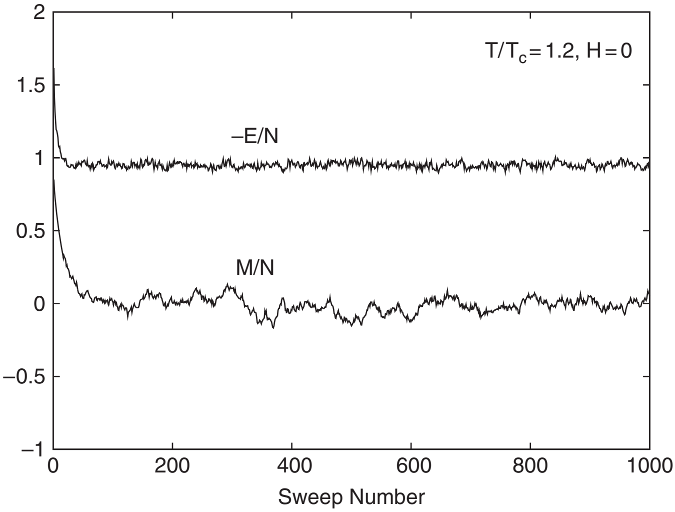

Figure 3.2 Relaxation to equilibrium as a function of the number of Monte Carlo sweeps for the magnetization per site M/N and the negative of the energy per site $-E/N$. The results are for the two-dimensional zero-field Ising model simulated by the heat-bath algorithm. $T = 1.2T_c$ and $H = 0$. The lattice size is 100 × 100.

for example, it would involve a Monte Carlo update of the Ising spin at each lattice site.

A simple way to estimate the number of equilibration steps is to plot the measured values as a function of the Monte Carlo sweep number and then conservatively estimate the number at which they begin to fluctuate around a mean value. We illustrate this technique in Fig. 3.2 where we present plots for the magnetization per site (M/N) and the negative of the energy per site ($-E/N$) for the two-dimensional zero field Ising model. We used the heat-bath algorithm and performed the simulation at $T = 1.2T_c$, that is, above the critical temperature in the disordered phase. The lattice size is 100 × 100, and our initial configuration was the state in which the Ising variable at each site was "up." From the figure it is clear that different physical quantities relax at different rates. The energy per site appears equilibrated within 50 sweeps. The equilibration time for the magnetization is longer and more ambiguous. A value between 500 and 1000 seems reasonable. Taking the larger value is wise.

While not presented here, it is useful to test the sensitivity of the estimate to changes in the random number seed. In many cases, it is important to try initial configurations with different symmetries. Also important is performing simulations with a number of sweeps much longer than the initially estimated equilibration time, plotting only a subset of values if convenient and appropriate, to check

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

whether the equilibrium plateau is stable, that is, whether it does not drift up or down or whether it does not eventually drop or rise precipitously.

Obviously, successive measurements shown in Fig. 3.2 are correlated. A proper definition of the equilibration time requires the calculation, for each physical quantity measured, of the *autocorrelation time*, which tells us over how many Monte Carlo sweeps the correlations persist. We give a precise definition of this quantity in Section 3.3. The *equilibration time* may then be defined as the maximum of the autocorrelation times for all observables of interest. In practice, we should choose the number of equilibration steps several times larger than this equilibration time.

3.2 Calculating averages and estimating errors

With the system equilibrated, we want to make measurements and estimate their statistical errors. Here we introduce the statistical basis for the common procedures.

The Monte Carlo sampling produces a sequence of configurations C_i from which we calculate a sequence of values $x_i = X(C_i)$ of some physical observable X. A function of a random variable is also a random variable. Hence, for each observable, we produce a different sequence of the values of a random variable. The properties of a random variable X follow from its distribution function $f_X(x)$. In Monte Carlo simulations we are mainly interested in two properties, the *mean*

$$\langle x \rangle = \int dx\, x f_X(x), \tag{3.1}$$

which is also called the *average* or the *expectation value*, and the *variance*

$$\sigma_X^2 = \int dx (x - \langle x \rangle)^2 f_X(x) = \langle x^2 \rangle - \langle x \rangle^2, \tag{3.2}$$

which is also called the *dispersion*.

With the mean and variance known, *Tchebycheff's inequality* allows us to calculate the likelihood that the value of the random variable lies in a certain interval. This inequality says that *independent of the form of* $f_X(x)$

$$P\left(\langle x \rangle - k\sigma_X < x < \langle x \rangle + k\sigma_X\right) \geq 1 - \frac{1}{k^2}. \tag{3.3}$$

This theorem is relatively easy to prove: From the definition (3.2) of the variance, it follows that

$$\begin{aligned}\sigma_X^2 \geq \int_{-\infty}^{\langle x \rangle - k\sigma} (x - \langle x \rangle)^2 f_X(x) dx + \int_{\langle x \rangle + k\sigma}^{\infty} (x - \langle x \rangle)^2 f_X(x) dx \\ \geq (k\sigma)^2 \int_{-\infty}^{\langle x \rangle - k\sigma} f_X(x) dx + (k\sigma)^2 \int_{\langle x \rangle + k\sigma}^{\infty} f_X(x) dx.\end{aligned}$$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

Hence, upon setting $\sigma^2 = \sigma_X^2$, we obtain $P(|x - \langle x\rangle| \geq k\sigma_X) \quad \leq \quad k^{-2}$, or equivalently (3.3).

Even in a detailed balance simulation, where we know a priori the distribution function of the configuration, we do not know a priori the mean and variance for any physical observable we are measuring relative to this distribution, so we cannot yet use Tchebycheff's inequality.

To obtain this information we begin with a sequence $x_1, x_2, \ldots, x_M$ of values of the random variable X, which we assume is uncorrelated. Their arithmetic mean

$$\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{3.4}$$

is called the *sample mean*. The expectation value of the sample mean is

$$\langle \bar{x}\rangle = \frac{1}{M}\sum_{i=1}^{M}\langle x_i\rangle = \frac{1}{M}\sum_{i=1}^{M}\langle x\rangle = \langle x\rangle$$

and its variance is

$$\begin{aligned}\sigma_{\bar{X}}^2 = \left\langle \bar{x}^2\right\rangle - \langle \bar{x}\rangle^2 &= \frac{1}{M^2}\left\langle \sum_{i,j} x_i x_j\right\rangle - \langle x\rangle^2 \\ &= \frac{1}{M^2}\left\langle \sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right\rangle - \langle x\rangle^2. \end{aligned}\tag{3.5}$$

The absence of correlations among the x_i implies $\langle x_i x_j\rangle = \langle x_i\rangle\langle x_j\rangle = \langle x\rangle^2$, so the above becomes

$$\sigma_{\bar{X}}^2 = \frac{1}{M}\left\langle \frac{1}{M}\sum_{i=1}^{M} x_i^2\right\rangle - \frac{1}{M}\langle x\rangle^2 = \frac{\sigma_X^2}{M}. \tag{3.6}$$

Thus, we have shown that the sample mean is the same as the mean of the random variable, but its variance is reduced by the number of terms in the sequence. Tchebycheff's inequality implies the confidence interval

$$P\left(\langle x\rangle - k\frac{\sigma_X}{\sqrt{M}} < \bar{x} < \langle x\rangle + k\frac{\sigma_X}{\sqrt{M}}\right) \geq 1 - \frac{1}{k^2}, \tag{3.7}$$

or equivalently that

$$P\left(\bar{x} - k\frac{\sigma_X}{\sqrt{M}} < \langle x\rangle < \bar{x} + k\frac{\sigma_X}{\sqrt{M}}\right) \geq 1 - \frac{1}{k^2}.$$

We see that as M becomes large, $\bar{x} \rightarrow \langle x\rangle$.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

Our discussion assumed we know the variance σ_X^2. We do not. How do we estimate it? Corresponding to the sample mean is a *sample variance*,

$$s_{\bar{X}}^2 = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \bar{x})^2. \tag{3.8}$$

$s_{\bar{X}}$ is called the *standard deviation*. We previously showed that $\langle \bar{x} \rangle = \langle x \rangle$ and $\sigma_{\bar{X}}^2 = \sigma_X^2/M$. We now show that

$$\left\langle s_{\bar{X}}^2 \right\rangle = \sigma_X^2. \tag{3.9}$$

First we write $(x_i - \langle x \rangle)^2 = (x_i - \bar{x})^2 + 2\,(x_i - \bar{x})\,(\bar{x} - \langle x \rangle) + (\bar{x} - \langle x \rangle)^2$. Summing both sides of this equation from 1 to M, observing that $\sum_{i=1}^{M}(x_i - \bar{x}) = 0$, and taking the expectation value, we find that

$$M\sigma_X^2 = \left\langle \sum_{i=1}^{M}(x_i - \bar{x})^2 \right\rangle + M\sigma_{\bar{X}}^2 = \left\langle \sum_{i=1}^{M}(x_i - \bar{x})^2 \right\rangle + \sigma_X^2,$$

from which (3.9) follows.

The *central limit theorem* is a stronger theorem than Tchebycheff's inequality. It enables us to define tighter confidence intervals and to estimate a minimal number of measurements to obtain this confidence. Under very general conditions, mainly that the variance of the distribution is finite, this theorem says that as $M \to \infty$ the distribution of the random variable $\bar{x} = \sum_i x_i/M$ approaches a normal distribution

$$f_{\bar{X}}(\bar{x}) = \frac{1}{\sqrt{2\pi\sigma_{\bar{X}}^2}} e^{-(\langle x \rangle - \bar{x})^2/2\sigma_{\bar{X}}^2}.$$

With the theorem telling us that the form of the limiting distribution is Gaussian, we are able to use the well-known properties of this function to state tighter confidence intervals,

$$P\left(\bar{x} - \sigma_{\bar{X}} < \langle x \rangle < \bar{x} + \sigma_{\bar{X}}\right) = 0.65, \tag{3.10}$$

$$P\left(\bar{x} - 2\sigma_{\bar{X}} < \langle x \rangle < \bar{x} + 2\sigma_{\bar{X}}\right) = 0.95, \tag{3.11}$$

$$P\left(\bar{x} - 3\sigma_{\bar{X}} < \langle x \rangle < \bar{x} + 3\sigma_{\bar{X}}\right) = 0.99, \tag{3.12}$$

which are known as the "one-sigma," "two-sigma," and "three-sigma" confidence intervals.

The analysis of this section was based on the assumption that the Monte Carlo configurations are uncorrelated. In practical simulations, this is not the case, and we therefore must shift our attention to promoting statistical independence so we can use the expressions for the sample variance and the confidence intervals (3.10)–(3.12) to estimate error bars.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

3.3 Correlated measurements and autocorrelation times

If we retrace the analysis of the last section, we can easily convince ourselves that correlations in a sequence of measured values affect only our estimation of the variance but do not affect our estimation of the mean. Let us return to our calculation of the variance of $\bar{x}$, (3.5), and proceed without the assumption of statistical independence:

$$\begin{aligned}\sigma_{\bar{X}}^2 = \left\langle \bar{x}^2 \right\rangle - \langle \bar{x} \rangle^2 &= \frac{1}{M^2} \left\langle \sum_{ij} (x_i x_j - \langle x_i \rangle \langle x_j \rangle) \right\rangle \\ &= \frac{1}{M^2} \left\langle \sum_i \left(x_i^2 - \langle x_i \rangle^2 \right) + 2 \sum_{j>i} \left(x_i x_j - \langle x_i \rangle \langle x_j \rangle \right) \right\rangle . \end{aligned} \tag{3.13}$$

Defining the correlation function

$$\chi_{|i-j|} \equiv \chi_{ij} \equiv \left\langle x_i x_j \right\rangle - \langle x_i \rangle \left\langle x_j \right\rangle$$

we first remark that it is a function of only $|i-j|$ and $\chi_0 = \langle x^2 \rangle - \langle x \rangle^2 = \sigma_X^2$. Thus, we can rewrite (3.13) as

$$\sigma_{\bar{X}}^2 = \frac{1}{M} \left[\sigma_X^2 + 2 \sum_{k=1}^{M-1} \left(1 - \frac{k}{M} \right) \chi_k \right],$$

which in turn we rewrite as

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{M} \left(1 + 2\tau_X \right), \tag{3.14}$$

where

$$\tau_X = \sum_{k=1}^{M-1} \left(1 - \frac{k}{M} \right) \frac{\chi_k}{\chi_0}. \tag{3.15}$$

Inspection shows that $\tau_X \geq 0$ and equals zero only if the data are uncorrelated, that is, only if $\chi_k = 0$ for $k \neq 0$. If so, (3.14) becomes our previous result (3.6). If correlations are present, (3.14) says that for a fixed number of measurements M, the variance of $\bar{x}$ increases. Correlations reduce the effective amount of statistical information, since we generated only $M/(1 + 2\tau_X)$ uncorrelated samples. If we were to ignore the correlations, (3.14) implies that our error estimates would be too small. This discrepancy underscores the need for care in estimating the variance.

The quantity τ_X defined in (3.15) is called the *autocorrelation time* for the observable X. Computing σ_X^2 and τ_X is one way of estimating $\sigma_{\bar{X}}^2$. The problem is finding an unbiased estimator of τ_X. The discussion in the following section suggests another procedure, based on promoting the statistical independence of

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

measurements, that goes by the names of binning, batching, bunching, blocking, etc. We adopt the terminology of *blocking*.

3.4 Blocking analysis

The basis for the blocking method is the following. Let us first define the function $\mathrm{var}(X) = \sigma_X^2$, which returns the variance of a random variable X. We note that

$$\mathrm{var}(aX) = a^2\,\mathrm{var}(X), \tag{3.16}$$

where a is some constant and

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y), \tag{3.17}$$

if X and Y are independent. The sample mean is a random variable so we write its variance as

$$\mathrm{var}(\bar{X}) = \mathrm{var}\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right) = \mathrm{var}\left(\frac{1}{N_{\mathrm{blocks}}}\sum_{j=1}^{N_{\mathrm{blocks}}}\left(\frac{1}{m}\sum_{i=(j-1)m+1}^{jm} x_i\right)\right),$$

where in the last step we have split the M measurements into N_{blocks} blocks of length $m = M/N_{\mathrm{blocks}}$. We assume here that the block length is sufficiently long that the *block averages*

$$\bar{x}_j = \frac{1}{m}\sum_{i=(j-1)m+1}^{jm} x_i$$

are statistically independent. In this case, (3.16) and (3.17) imply

$$\mathrm{var}(\bar{X}) = \frac{1}{N_{\mathrm{blocks}}^2}\sum_{j=1}^{N_{\mathrm{blocks}}}\mathrm{var}(\bar{x}_j).$$

Next we assume that for sufficiently long blocks the block averages have a common variance. Thus,

$$\mathrm{var}(\bar{X}) = \frac{1}{N_{\mathrm{blocks}}}\mathrm{var}(\bar{X}_{\mathrm{block}}), \tag{3.18}$$

where $\bar{X}_{\mathrm{block}}$ is the random variable representing the block averages. Explicitly, we use the estimate

$$\mathrm{var}(\bar{X}_{\mathrm{block}}) \approx s^2_{\bar{X}_{\mathrm{block}}} = \frac{1}{N_{\mathrm{blocks}}-1}\sum_{j=1}^{N_{\mathrm{blocks}}}(\bar{x}_j - \bar{x})^2. \tag{3.19}$$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

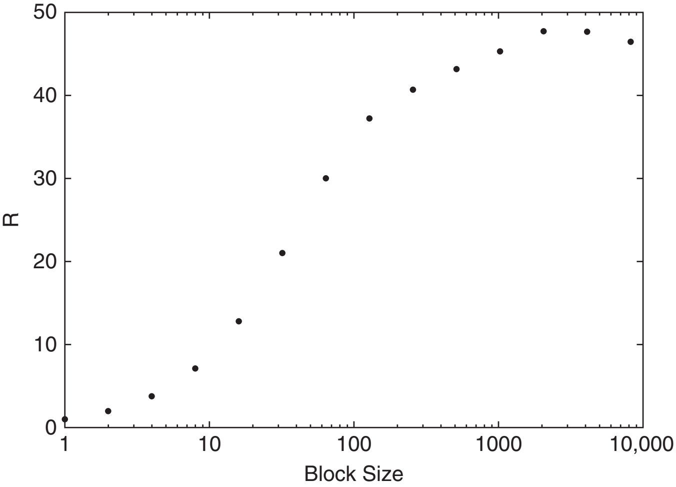

Figure 3.3 Estimating statistical independence by blocking. The simulation method, model, and its parameters are the same as in Fig. 3.2, and the observable is the magnetization per site. The size of the data stream was 1,000,000 with the first 500 eliminated. We started with a block size of $m = 1$ and successively doubled it.

The "one-sigma" error bar is given by $\sqrt{\mathrm{var}(\bar{X})}$. To determine $\mathrm{var}(\bar{X})$, we compute

$$R_X \equiv \frac{m\,\mathrm{var}(\bar{X}_{\mathrm{block}})}{\mathrm{var}(X)},$$

which *in the limit of independent blocks* relates $\mathrm{var}(\bar{X})$ to the "uncorrelated" estimate $\mathrm{var}(X)/M$:[1]

$$\mathrm{var}(\bar{X}) = R_X \frac{\mathrm{var}(X)}{M}. \qquad (3.20)$$

We plot R_X for increasing block sizes m, as shown in Fig. 3.3 for the magnetization per site (the full data stream's length is $M = 10^6$). As the block size increases, the block averages become statistically independent, and $\mathrm{var}(\bar{X}_{\mathrm{block}})$ becomes inversely propositional to m. When this happens, R_X saturates and we can then use (3.18) to obtain $\mathrm{var}(\bar{X})$. In the example of Fig. 3.3, saturation occurs around $m^{\mathrm{plateau}} = 1024$ and the saturation value is $R_X^{\mathrm{plateau}} \approx 50$. This means that by neglecting correlations between successive measurements, we would have underestimated the one-sigma error bar by a factor $\sqrt{50} \approx 7$. If no plateau is reached before the block size m

[1] Using (3.18) one finds $R\frac{\mathrm{var}(X)}{M} = \frac{m}{M}\mathrm{var}(\bar{X}_{\mathrm{blocks}}) = \frac{\mathrm{var}(\bar{X}_{\mathrm{blocks}})}{N_{\mathrm{blocks}}} = \mathrm{var}(\bar{X})$.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

approaches the largest useful value of about $\mathcal{O}(M/100)$, then we need much more data to compute a meaningful error bar.

From the definitions (3.14) and (3.20) it follows that the autocorrelation time τ is related to the value R_X^{plateau} by

$$\tau_X = \frac{1}{2}\left(R_X^{\text{plateau}} - 1\right). \tag{3.21}$$

The autocorrelation times for different observables differ. The largest value determines the equilibration time.

3.5 Data sufficiency

The central limit theorem does not tell us how many independent blocks we need, so the distribution of our data becomes Gaussian. It turns out N_{blocks} need not be that large. When one of the authors was writing his first Monte Carlo code, he asked a senior colleague who had been performing Monte Carlo simulations for several decades, "How many uncorrelated values do I need?" He replied, "32," and walked away. This brusque reply is usually about right.

While the field of statistics has many procedures for defining the probability that a set of data is distributed as a Gaussian, a convenient, simple, and useful alternative is overlaying a histogram of the data with a Gaussian centered at $\bar{x}$ and half-width $\sigma_{\bar{X}}$ to see if they look alike. We illustrate this in Fig. 3.4 for the cases of $N_{\text{blocks}} = 32$ and 64. In both cases, the agreement is actually satisfactory, even though the "fit" might look poor. In each case we have 20 bins in our histogram. Throwing 32 or 64 values into these bins produces histograms with bin counts subject to large fluctuations relative to the number of data in a bin.[2]

Deviations from Gaussian behavior tend to appear as a histogram leaning too left or too right. This is called *skewness*. Or the histogram appears too narrow or too squat. This is called *kurtosis*. Multiple peaks can also occur. A Gaussian has a skewness and a kurtosis of 0.[3] As a supplementary test quantitative measures of both quantities,

$$\text{skewness} = \frac{\sum_{i=1}^{N_{\text{blocks}}} (\bar{x}_i - \bar{x})^3}{(N_{\text{blocks}} - 1)\, s_{\bar{X}}^3}, \quad \text{kurtosis} = \frac{\sum_{i=1}^{N_{\text{blocks}}} (\bar{x}_i - \bar{x})^4}{(N_{\text{blocks}} - 1)\, s_{\bar{X}}^4} - 3$$

[2] We constructed our histograms in the following manner: A Gaussian, centered around 0 with a variance of 1, has an effective width of 6 ranging from -3 to 3 along the x-axis. For the case of data with an arbitrary mean and variance, we chose our bin width Δ to be $\Delta = 6\sigma_{\bar{X}}^2/20$ and positioned our bins so that one was centered at $\bar{x}$. Instead of a normalized Gaussian $G(\bar{x}, \sigma_{\bar{X}})$, we plotted $N_{\text{blocks}} G(\bar{x}, \sigma_{\bar{X}})$, that is, the expected count number in the interval $(x, x + \Delta)$.

[3] The mathematical definition of kurtosis yields 3 for a Gaussian. For convenience, one often subtracts 3 from the definition, so that Gaussian behavior corresponds to both the skewness and kurtosis approaching 0.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

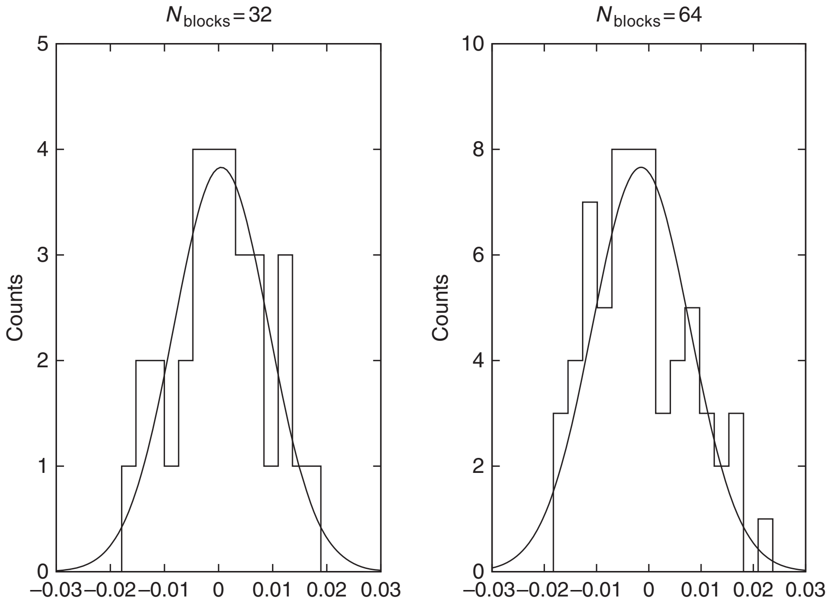

Figure 3.4 Histograms and expected Gaussian counts. The simulation method, model, and its parameters are the same as in Fig. 3.2.

are easily computed for comparison to the Gaussian values. Still another test is a goodness-to-fit test between the Gaussian and the histogram.

A *control chart* is a simple, quick, visual check of the consistency of the proposed values of equilibration steps, block size, and sweeps. Plotted in Fig. 3.5 are the block averages of the energy as a function of block number. The solid horizontal line is the average magnetization per site; the dashed horizontal lines lie one and two sigmas above and below it. We used the proverbial 32 blocks. If we underestimated the number of equilibration steps badly, we would see a drift in the data. This appears not to be the case. If the number of sweeps and the block size are about right, we expect to see about 21 (0.65 times 32) markers between $\bar{x} - \sigma_{\bar{X}}$ and $\bar{x} + \sigma_{\bar{X}}$, about 30 or 31 (0.95 times 32), between $\bar{x} - 2\sigma_{\bar{X}}$ and $\bar{x} + 2\sigma_{\bar{X}}$, and hence maybe 1 or 2 outside this range. Our control chart is consistent with these expectations. Because these confidence intervals are statistically based, fluctuations might produce deviations from the expected values. This type of chart provides a good test for whether the data require a closer analysis.

It is also useful to count in this chart the number of "runs up" and "runs down," that is, the number of successive values that increase or decrease. If the data were Gaussian distributed, the result would look like "white noise" with small runs up and down. If the runs are consistently large, then the block averages may be correlated and further analysis is warranted.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

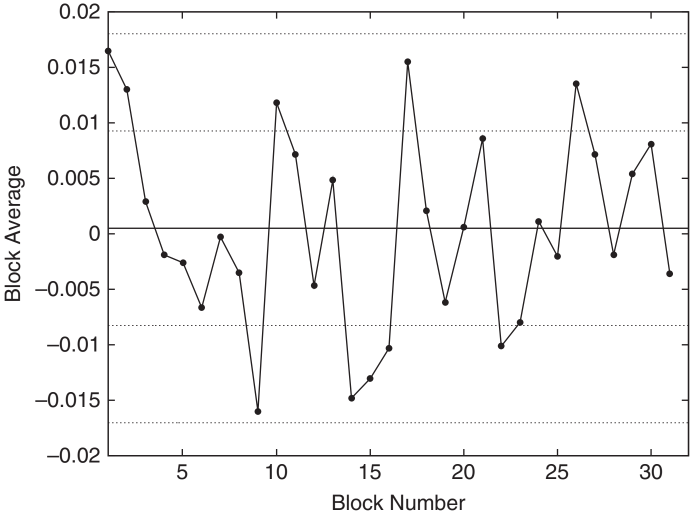

Figure 3.5 Control chart for the $N_{\text{blocks}} = 32$ case of Fig. 3.4.

While the data analysis discussed above provides an estimate of the error, a given application might require the error bars to be small. To make them smaller we generate more blocks. The error decreases as $1/\sqrt{N_{\text{blocks}}}$, while the computation time increases proportional to N_{blocks}.

The field of statistics has numerous additional tests and theorems that are sometimes helpful for troublesome simulations. They lie outside the scope of this book. The methods discussed in this chapter are generally sufficient for analyzing Monte Carlo data and stating with confidence the estimated means and their statistical errors. There is another issue that should be kept in mind: Unless the simulation is properly equilibrated and the sampling is ergodic, these estimates might be meaningless. As noted in Section 2.4.2, a simulation can get stuck in a limited region of phase space. The mean values and errors of the data sampled then reflect only this part of phase space and might be completely unrelated to the values obtained if the sampling visited the entire phase with the proper frequency.

We began this chapter by discussing the equilibration of the sampling, noting certain “sticking” mechanisms, which might affect the simulations of both first-order and continuous phase transitions. We also illustrated (Fig. 3.2) that two different observables generally equilibrate at different rates. Somewhat in passing, we noted the need to experiment with the simulation to test the sensitivity of the results to changes in initial configurations, their symmetries, etc. These studies and others are just as important for the uncertainty quantification of the simulation as the relatively routine statistical analysis just discussed. Different problems generally require different types of experimentation. There are no sure-fire procedures to

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

ensure equilibration and ergodic sampling. The experience of the researcher and the use of effective algorithms are important factors.

3.6 Error propagation

Sometimes we need to estimate the variance of a function of two or more random variables. An example of such a quantity of interest is the *Binder ratio*, defined as

$$B_X = \frac{\langle X^4 \rangle}{\langle X^2 \rangle^2}. \tag{3.22}$$

With the random variable X being the magnetization, this ratio is useful in the finite-size scaling analysis of phase boundaries of the Ising model and other spin models. Because the values of the numerator and denominator are computed with the same $\{x_i = X(C_i)\}$, they are statistically correlated. In computing the error associated with B_X, how do we account for these correlations?

In making multivariate variance estimates, several generalizations of the concept of variance are useful. The *covariance* of two random variables X and Y is

$$\mathrm{cov}(X,Y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle, \tag{3.23}$$

and the *correlation coefficient* of these two variables is

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

The numerical estimate of $\mathrm{cov}(X,Y)$ is

$$\mathrm{cov}(X,Y) \approx \frac{1}{M-1} \sum_{i=1}^{M} (x_i - \bar{x})\,(y_i - \bar{y}), \tag{3.24}$$

and the numerical estimate of $\rho(X,Y)$ is

$$\rho(X,Y) \approx \frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{M} (x_i - \bar{x})^2 \sum_{j=1}^{M} (y_j - \bar{y})^2}}. \tag{3.25}$$

The correlation coefficient satisfies $|\rho(X,Y)| \le 1$.

Let us first discuss how to estimate $g(\langle x \rangle)$ for an arbitrary function g, and the associated error. One possibility to estimate $g(\langle x \rangle)$ is to compute

$$\overline{g(x)} \equiv \frac{1}{M} \sum_{i=1}^{M} g(x_i), \tag{3.26}$$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

and another is

$$g(\bar{x}) \equiv g\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right). \tag{3.27}$$

When M is large and g is nonlinear, the latter estimator is better. Expanding (3.26) with respect to $\Delta x \equiv x - \langle x \rangle$ and then taking the expectation value generates

$$\begin{aligned}\left\langle \overline{g(x)} \right\rangle &= \left\langle g(\langle x \rangle) + g'|_{\langle x \rangle}\Delta x + \frac{1}{2}g''|_{\langle x \rangle}(\Delta x)^2 + \cdots \right\rangle \\ &= g(\langle x \rangle) + \frac{1}{2}g''|_{\langle x \rangle}\mathrm{var}(X) + \cdots .\end{aligned}$$

Similarly, by expanding (3.27) with respect to $\Delta\bar{x} \equiv \bar{x} - \langle x \rangle$, we obtain

$$\begin{aligned}\langle g(\bar{x}) \rangle &= \left\langle g(\langle x \rangle) + g'|_{\langle x \rangle}\Delta\bar{x} + \frac{1}{2}g''|_{\langle x \rangle}(\Delta\bar{x})^2 + \cdots \right\rangle \\ &= g(\langle x \rangle) + \frac{1}{2}g''|_{\langle x \rangle}\mathrm{var}(\bar{X}) + \cdots ,\end{aligned} \tag{3.28}$$

where in the second term we have $\mathrm{var}(\bar{X}) = \mathrm{var}(X)/M$. We note that neither (3.26) nor (3.27) is an unbiased estimator for $g(\langle x \rangle)$. An estimator is *unbiased* if its expected value equals its true value. The above estimators have a systematic error proportional to the curvature of the function $g(x)$ at $x = \langle x \rangle$. However, for the second estimator (3.27), the systematic error is proportional to $1/M$. This bias is negligible because it is much smaller than the statistical error, which is proportional to $1/\sqrt{M}$.

In fact, let us calculate the statistical error of the estimator (3.27). The expectation value of the squared error is

$$\left\langle (\Delta g(\bar{x}))^2 \right\rangle = \left\langle (g(\bar{x}) - g(\langle x \rangle))^2 \right\rangle \approx \left\langle \left(g'|_{\bar{x}}\Delta\bar{x}\right)^2 \right\rangle = \left(g'|_{\bar{x}}\right)^2 \mathrm{var}(\bar{X}). \tag{3.29}$$

The numerical variance estimate is given by (3.18) and (3.19), and we evaluate g' at $\bar{x}$ instead of $\langle x \rangle$. This approximation should have a small effect on the estimate of the error.

Similarly, in the case of a bivariate function, we use the estimate $g(\langle x \rangle, \langle y \rangle) \approx g(\bar{x}, \bar{y})$ and compute the variance as

$$\begin{aligned}\left\langle (\Delta g(\bar{x}, \bar{y}))^2 \right\rangle &= \left\langle (g(\bar{x}, \bar{y}) - g(\langle \bar{x} \rangle, \langle \bar{y} \rangle))^2 \right\rangle \\ &\approx \left\langle \left(\partial_x g|_{\langle x \rangle, \langle y \rangle}\Delta\bar{x} + \partial_y g|_{\langle x \rangle, \langle y \rangle}\Delta\bar{y}\right)^2 \right\rangle \\ &\approx \left(\partial_x g|_{\bar{x},\bar{y}}\right)^2 \mathrm{var}(\bar{X}) + \left(\partial_y g|_{\bar{x},\bar{y}}\right)^2 \mathrm{var}(\bar{Y}) + 2\partial_x g|_{\bar{x},\bar{y}}\,\partial_y g|_{\bar{x},\bar{y}}\,\mathrm{cov}(\bar{X}, \bar{Y}),\end{aligned}$$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

where again we evaluate the derivatives at $x = \bar{x}$ and $y = \bar{y}$. For example, if $g(\langle x\rangle, \langle y\rangle) = \langle x\rangle/\langle y\rangle$, we find

$$\frac{\left\langle(\Delta(\bar{x}/\bar{y}))^2\right\rangle}{(\bar{x}/\bar{y})^2} \approx \frac{\operatorname{var}(\bar{X})}{\bar{x}^2} + \frac{\operatorname{var}(\bar{Y})}{\bar{y}^2} - 2\frac{\operatorname{cov}(\bar{X}, \bar{Y})}{\bar{x}\bar{y}}. \tag{3.30}$$

This particular expression is useful for error estimations, for example, when the simulation has a small to moderate sign problem (Section 5.4). In this case, the numerator is the sign-weighted average and the denominator is the average sign.

3.7 Jackknife analysis

It can be tedious to compute the whole correlation matrix and derivatives of a function to estimate the function of a mean and its variance. The jackknife method is a data resampling method which often provides a good alternative. Its use is simple in the sense that the estimates are generalizations of the standard definitions of a sample mean (3.4) and a sample variance (3.8).

Let us define the "deleted average" $x_{[i]}$ as the sample average in which we ignore the value x_i,

$$x_{[i]} = \frac{1}{M-1}\sum_{j\neq i} x_j = \frac{M\bar{x} - x_i}{M-1}.$$

It is a simple matter to show that the sample average of the deleted averages $x_{[i]}$ equals the sample average $\bar{x}$,

$$\overline{x_{[\,]}} = \bar{x}.$$

It is also a simple matter to show that

$$s^2_{[\,]} = \frac{M-1}{M}\sum_{i=1}^{M}\left(x_{[i]} - \overline{x_{[\,]}}\right)^2$$

equals the sample variance (3.8).

In practice, the jackknife method is applied to the block averages that we introduced in the discussion of the blocking analysis. The procedure is as follows: We split the M measurements, which should be large, into N_{blocks} blocks of length m, with m larger than the autocorrelation time τ. Then we compute the averages

$$x_{[i]} = \frac{1}{N_{\text{blocks}} - 1}\sum_{j\neq i} \bar{x}_j, \quad i = 1, \ldots, N_{\text{blocks}}, \tag{3.31}$$

where we use all the blocks, *except* block number i, and $\bar{x}_j$ denotes the average over block j. We next define $g_{[i]} = g(x_{[i]})$. The simple jackknife estimate of $g(\langle x\rangle)$ is the average $\overline{g_{[\,]}}$ of these $g_{[i]}$:

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

$$g(\langle x\rangle) \approx \overline{g_{[\,]}} \equiv \frac{1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} g_{[i]}. \tag{3.32}$$

Using (3.19) and (3.18), we find that the expectation value of the Taylor series expansion of

$$g_{[i]} = g\left(\frac{1}{N_{\text{blocks}}-1} \sum_{j\neq i} \bar{x}_j\right) = g\left(\bar{x} + \frac{1}{N_{\text{blocks}}-1}(\bar{x}-\bar{x}_i)\right)$$

equals

$$\begin{aligned}\left\langle g_{[i]}\right\rangle &= \langle g(\bar{x})\rangle + \frac{1}{2(N_{\text{blocks}}-1)} g''|_{\bar{x}} \frac{\operatorname{var}(\bar{X}_{\text{block}})}{N_{\text{blocks}}} + \cdots \\ &= \langle g(\bar{x})\rangle + \frac{1}{2(N_{\text{blocks}}-1)} g''|_{\bar{x}} \operatorname{var}(\bar{X}) + \cdots .\end{aligned} \tag{3.33}$$

This equation says that the bias in the jackknife estimate is reduced compared with that in (3.28) by a factor $1/(N_{\text{blocks}}-1)$. By combining (3.28) and (3.33), we obtain an estimator without g''-bias:[4]

$$g(\langle x\rangle) \approx N_{\text{blocks}} g(\bar{x}) - (N_{\text{blocks}}-1)\overline{g_{[\,]}}. \tag{3.34}$$

We now show that the error estimate on $g(\langle x\rangle)$ is

$$\left\langle (\Delta g(\bar{x}))^2\right\rangle \approx \frac{N_{\text{blocks}}-1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} (g_{[i]} - g(\bar{x}))^2. \tag{3.35}$$

Since $g_{[i]} - g(\bar{x}) \approx g'|_{\bar{x}}(\bar{x}-\bar{x}_i)/(N_{\text{blocks}}-1)$, it follows from (3.18) and (3.19) that

$$\begin{aligned}\frac{N_{\text{blocks}}-1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} (g_{[i]} - g(\bar{x}))^2 &\approx \frac{(g'|_{\bar{x}})^2}{N_{\text{blocks}}(N_{\text{blocks}}-1)} \sum_{i=1}^{N_{\text{blocks}}} (\bar{x}-\bar{x}_i)^2 \\ &= (g'|_{\bar{x}})^2 \frac{\operatorname{var}(\bar{X}_{\text{blocks}})}{N_{\text{blocks}}} = (g'|_{\bar{x}})^2 \operatorname{var}(\bar{X}),\end{aligned}$$

which is consistent with (3.29).

The jackknife procedure is easily extended to arbitrary functions f of one or several observables, such as (3.22). If $f(\{\bar{x}_j\})$ is the estimate based on the expectation

[4] The expectation value of the right-hand side is $N_{\text{blocks}}\langle g(\bar{x})\rangle - (N_{\text{blocks}}-1)\langle\overline{g_{[\,]}}\rangle = \langle g(\bar{x})\rangle - \frac{1}{2}g''|_{\bar{x}} \operatorname{var}(\bar{X}) + \cdots = g(\langle x\rangle) + \frac{1}{2}g''|_{\bar{x}} \operatorname{var}(\bar{X}) - \frac{1}{2}g''|_{\bar{x}} \operatorname{var}(\bar{X}) + \cdots .$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

values of the observables in the block ensemble $\{\bar{x}_j\}$, we define $f_{[i]} = f(\{\bar{x}_{j\neq i}\})$. The jackknife estimate of the mean and variance is then

$$\overline{f_{[\,]}} = \frac{1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} f_{[i]}, \tag{3.36}$$

$$\langle (\Delta f)^2 \rangle = \frac{N_{\text{blocks}} - 1}{N_{\text{blocks}}} \sum_{i=1}^{N_{\text{blocks}}} (f_{[i]} - \overline{f_{[\,]}})^2. \tag{3.37}$$

Let us close this section with some remarks about the limitations of the jackknife method. By rewriting the definition (3.31) of the deleted averages as $x_{[i]} = \bar{x} + \frac{1}{N_{\text{blocks}}-1}(\bar{x} - \bar{x}_i)$, it becomes obvious that the members of the jackknife ensemble are distributed in a small sample space, centered around the average of the original ensemble, whose linear size is $N_{\text{blocks}} - 1$ times smaller than that of the original ensemble. Therefore, the linear size of this space is approximately $\sqrt{N_{\text{blocks}}}$ times smaller than the statistical error, which is $\mathcal{O}(1/\sqrt{N_{\text{blocks}}})$. As a result, we need to multiply the variance of the jackknife ensemble in (3.35) by a factor of order N_{blocks} to obtain the estimator for the squared statistical error in $g(\bar{x})$. While the jackknife ensemble picks up the local curvature of g correctly and works well when the gradient of the function does not change much in the region of size $1/\sqrt{N_{\text{blocks}}}$ centered around the correct expectation value, it may fail when the nonlinearity of the function g is strong (Fig. 3.6). In the latter case, the small sample space covered by the jackknife ensemble may not include the correct expectation value $\langle x \rangle$.

3.8 Bootstrap analysis

We can overcome the limitations of the jackknife analysis by employing a resampling method that mimics the broader distribution of the sample average $\bar{x}$. The *bootstrap method* is a popular choice for this purpose. A member of this ensemble, which we denote by x'_i $(i = 1, 2, \ldots, N_{\text{bootstrap}})$, is generated by randomly choosing N_{blocks} samples $\bar{x}_j$ from the original ensemble of block averages, without trying to avoid picking the same one multiple times. As a result, some members of the original ensemble are not chosen at all, while some are chosen once, some twice, and so on. Then x'_i is defined as the average of these N_{blocks} values. Thus, formally, $x'_i = \frac{1}{N_{\text{blocks}}} \sum_{k=1}^{N_{\text{blocks}}} \bar{x}_{\rho(i,k)}$, where $\rho(i,k) = 1, 2, \ldots, N_{\text{blocks}}$ is a uniform random integer such that $\rho(i,k)$ and $\rho(j,k')$ are statistically independent if $(i,k) \neq (j,k')$.

Because of this averaging, the standard deviation of the bootstrap ensemble is of the same order as the standard deviation of the original ensemble multiplied by $1/\sqrt{N_{\text{blocks}}}$, i.e., $\text{var}(\{x'_i\}) \sim \text{var}(\{\bar{x}_i\})/N_{\text{blocks}}$. Hence, the variance of the bootstrap

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

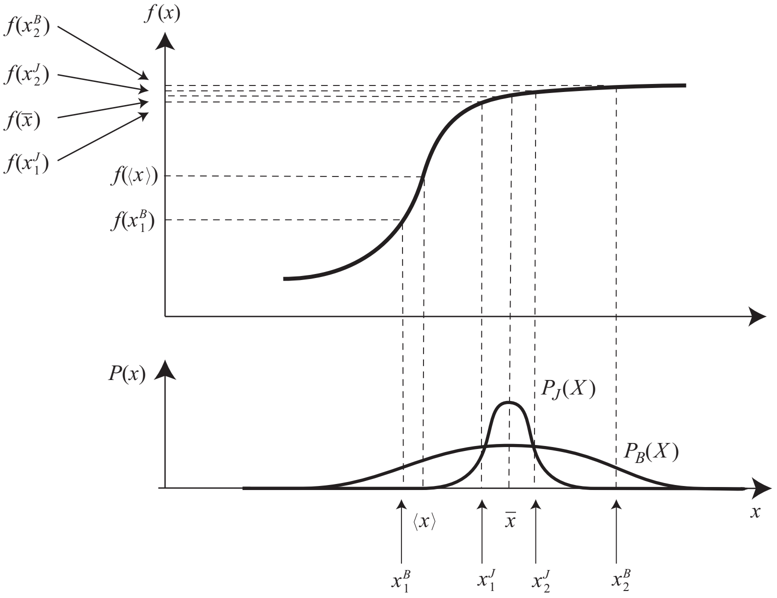

Figure 3.6 Estimation of a strongly nonlinear function by jackknife and bootstrap analyses. The bootstrap samples are distributed, roughly, over the region $x_1^B < x < x_2^B$. As a result, the confidence interval for $f^* \equiv f(\langle x\rangle)$ is $f(x_1^B) < f^* < f(x_2^B)$. The jackknife samples are distributed over a much narrower interval $x_1^J < x < x_2^J$. The confidence interval for f^* of the jackknife analysis is then $f(x_1^J) < f^* < f(x_2^J)$ expanded by the factor $\sqrt{N_{\text{blocks}}}$, which is about 4 in this example. Obviously, the factor 4 is not sufficient to cover the correct value of f^*.

ensemble is comparable to the variance of the sample average: $\text{var}(\{x_i'\}) \sim \text{var}(\bar{x})$. As a result, the correct value $\langle x\rangle$ should be contained with high probability in the sample space covered by the bootstrap ensemble, and the correct value of $g(\langle x\rangle)$ should fall within the image of this region. We illustrate this in Fig. 3.6.

In case the non-linearity is not so strong, a little more precise … arithmetic yields that the estimate of the statistical error in $g(\bar{x})$ is obtained by the variance of the bootstrap samples as

$$\sigma^2 \equiv \left\langle (g(\bar{x}) - g(\langle x\rangle))^2 \right\rangle \approx \frac{N_{\text{blocks}}}{N_{\text{blocks}} - 1} \left\langle \left(g(x_i') - g(\bar{x})\right)^2 \right\rangle$$

when $N_{\text{bootstrap}} \gg 1$. In contrast to the jackknife analysis, this estimator should produce a sensible result even in the case where the nonlinearity of the function g is strong.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

3.9 Monte Carlo computer program

Algorithm 5 shows the core structure of a Monte Carlo program, which is very simple. This structure assumes that the data analysis is a post-simulation process. Besides the continuing dramatic increases in processor speeds, the cost and access speeds of external memory have also improved dramatically over the past few years, making this separation of tasks feasible and advantageous. Not too many years ago it would not always have been possible. Some of the methods mentioned for the analysis of the data, such as using a control chart and the jackknife method, in fact require a post-simulation analysis.

Algorithm 5 requires as input the parameters of the physical model. It first defines or reads the simulation parameters, initializes arrays and variables needed by the program, defines the initial Monte Carlo configuration, and perhaps even sets the seed for the random number generator. The most important building block of the program is a routine that performs an update of the entire configuration. For example, in a simulation of the classical Ising model based on local updates, this function would loop over all lattice sites and propose an update for each spin. We call this a "Monte Carlo sweep." The second most important procedure is the one that computes observables, such as the energy or the magnetization, for a given Monte Carlo configuration. Which quantities are computed depends on the problem. As the measurements are made, they are accumulated in an appropriate data container. This data collection is the result of the simulation and is saved at the end of the simulation. Depending on the amount of data and the time needed to save the data relative to the execution cost of a sweep, the data might instead

Algorithm 5 Structure of a Monte Carlo program.

```
Input: Model and control parameters ($N_{\text{equil}}$ = number of equilibration sweeps,
  $N_{\text{sweep}}$ = number of sweeps, $N_{\text{skip}}$ = number of updates between measurements),
  plus accumulators for storing measured quantities.
  Initialize simulation ;
  for $i = 1$ to $N_{\text{equil}}$ do
    Perform Monte Carlo sweep ;
  end for
  for $i = 1$ to $N_{\text{sweep}}$ do
    Perform a Monte Carlo sweep ;
    if $i$ is multiple of $N_{\text{skip}}$ then
      Perform measurements and accumulate the results ;
    end if
  end for
  return the accumulators.
```

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

be written immediately after the measurement, thus eliminating the final output step. To avoid a bias from the choice of the initial configuration, the first N_{equil} sweeps (with N_{equil} larger than the thermalization time) are not measured. N_{sweep} is the number of sweeps performed after equilibration. The cost of computing the measurements relative to the cost of a sweep often makes it advisable to measure and save the measurements from only every N_{skip} configurations, with N_{skip} of the order of the shortest autocorrelation time.

For post-simulation data processing, another computer program loads the data from the disk and computes means and variances. For the blocking analysis, a loop over the block sizes $m_l = 2^l$ is performed, and the averages of the measured observables are computed for each block. The variance of these block-averaged measurements divided by the number of blocks yields the square of the error estimate for iteration l. The increase of the error with increasing l is monitored, and if saturation occurs, the saturated value gives a reliable estimate of the square of a one-sigma error bar. If no saturation occurs up to the largest possible block size (which is about two orders of magnitude smaller than the length of the data stream), then a reliable error estimate is not possible. In this case, much more simulation data are needed.

Some simulations require the computation of a lot of observables. While locating the phase boundary of the Ising model requires the computation of just a few simple quantities, such as the energy and the order parameter, the discovery of a proposed novel phase in a more exotic model might require the computation of the spatial dependence of multiple correlation functions for a wide range of model and control parameters. In such cases, the storage cost may become an issue, and the data file management requires pre-simulation thought.

Different people have different preferences about doing all or at least part of the data analysis on the fly instead of doing it all after the simulation. On a practical note, outputting a stream of on-the-fly block averages of select measurements is a convenient real-time indicator of the state of the simulation. Having such indicators is especially helpful in the code development phase, that is, in the development of the procedure that updates the configurations. In the code development phase, the ability to rerun a simulation with the same sequence of random numbers is also essential. Even if all the measurements are done on the fly, it is advisable to output to external storage a few streams of select measurements for inspection in case something weird appears to be happening. As a stochastic process, a Monte Carlo simulation is susceptible to rare events occurring. Having these streams of information on hand to help distinguish among rare events, coding errors, and algorithm instabilities might help to avoid rerunning the simulation (using the same random number seed) to replicate the problem and might help to resolve the issue more rapidly.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

Suggested reading

M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, vol. 1: *The Basics* (New York: Wiley-Interscience, 1986), chapter 2.

A. Papoulis, *Probability and Statistics* (Englewood Cliffs, NJ: Prentice Hall, 1990), chapters 4, 7, and 9.

W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, 1992), chapter 14.

Exercises

3.1 Write a short computer program to simulate the two-dimensional Ising model in a magnetic field, using either the heat-bath or Metropolis algorithm. Examples of FORTRAN and C programs of varying degrees of complexity can be found by searching the Web or by looking in Parisi (1988), Landau and Binder (2000), or Newman and Barkema (1999). Add the computation of the specific heat to the set of measurements.

3.2 For a 32 × 32 lattice, repeat the analysis in Figs. 3.2 through 3.5.

3.3 Repeat Exercise 3.2 for a series of temperatures approaching T_c.

3.4 Repeat Exercise 3.2 for a temperature equal to T_c for several large lattice sizes.

3.5 Repeat Exercise 3.2 for a series of magnetic fields increasing from zero.

3.6 If $\text{var}(X)$ is the function that returns the variance (3.2) of the random variable X and if $\text{cov}(X, Y)$ returns the covariance (3.23) of the random variables X and Y, prove that $\text{cov}(X, Y)^2 \leq \text{var}(X)\text{var}(Y)$ and hence that $-1 \leq \rho(X, Y) \leq 1$.

3.7 The *conditional mean* of a function $g(x, y)$ with respect to the variable x relative to a distribution $p(x, y)$ is

$$E\left[g(x, y) | \, x\right] = \int g(x, y) p(x, y) dy.$$

Show

1. $E\left[g(x, y) | \, x\right] = \int g(x, y) p(x, y) dy$.
2. $E\left[E\left[g(x, y) | \, x\right]\right] = E\left[g(x, y)\right]$.
3. $E\left[g_1(x) g_2(y) \, |x\right] = g_1(x) E\left[g_2(y) \, |x\right]$.
4. $E\left[g_1(x) g_2(y)\right] = E\left[g_1(x) E\left[g_2(y) \, |x\right]\right]$.

3.8 A *conditional variance* is defined as

$$\begin{aligned}\text{var}\left[g(x, y) \, |x\right] &= E\left[\left(g(x, y) - E\left[g(x, y) \, |x\right]\right)^2\right] \\ &= E\left[g^2(x, y) \, |x\right] - E^2\left[g(x, y) \, |x\right].\end{aligned}$$

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

Show that

$$E\left[\operatorname{var}\left[g(x,y|x)\right]\right]+\operatorname{var}\left[E\left[g(x,y)\,|x\right]\right]$$
$$=E\left[g^2(x,y)\right]-E^2\left[g(x,y)\right]=\operatorname{var}\left[g(x,y)\right].$$

3.9 For a target distribution $p(z)$, the Monte Carlo method estimates $E[g(z)]$ via

$$I_1=\frac{1}{M}\sum_{i=1}^{M}g\left(z_i\right),$$

where the z_i are samples drawn for $p(z)$. If we split up z as (x,y) and assume we know or can calculate easily $E[g(x,y)|x]$, then an alternative estimate of $E[g(z)]$ is

$$I_2=\frac{1}{M}\sum_{i=1}^{M}E\left[g(z)|x_i\right].$$

Use the results of the two previous exercises to argue that

1. Both estimates I_1 and I_2 are unbiased and give the same expectation value.
2. If the computational costs are comparable, then the second estimate I_2 is preferred because its variance is potentially smaller.

These results illustrate a Monte Carlo rule of thumb: If part of the sampling is replaced by an exact result, the variance is often reduced. Numerous variance reduction methods are based on this rule.

3.10 If $\{f_i(X_j)\}$ is a set of functions of a set $\{X_j\}$ of random variables X_j, show that

$$\operatorname{var}(f_i)=\sum_j\left(\frac{\partial f_i}{\partial x_j}\right)\operatorname{var}(X_j)$$
$$+\sum_j\sum_{j\neq k}\left(\frac{\partial f_i}{\partial x_j}\right)\left(\frac{\partial f_i}{\partial x_k}\right)\operatorname{cov}(X_j,X_k),$$
$$\operatorname{cov}(f_i,f_j)=\sum_{kl}\left(\frac{\partial f_i}{\partial x_k}\right)\left(\frac{\partial f_j}{\partial x_k}\right)\operatorname{var}(X_k)$$
$$+\sum_k\sum_{k\neq l}\left(\frac{\partial f_i}{\partial x_k}\right)\left(\frac{\partial f_j}{\partial x_l}\right)\operatorname{cov}(X_k,X_l).$$

The functions $\operatorname{var}(X)$ and $\operatorname{cov}(X,Y)$ are defined in Exercise 3.6.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press

3.11 In a one-dimensional random walk along a line, the walker at the i-th position makes a displacement of $X_i = \pm d$ from its current position with equal probability. After n steps, its displacement is $X(n) = X_1 + X_2 + X_n = \cdots + X_n$.

1. Show that $\langle X(n) \rangle = 0$ and $\langle X(n)^2 \rangle = nd^d$.
2. Find $\mathrm{cov}(X(n), X(m))$ and $\rho(X(n), Y(n))$.
3. Show that the $X(m)$ and $X(n)$ become completely correlated as $m/n \to 1$ and completely uncorrelated as $m/n \to 0$.

https://doi.org/10.1017/CBO9780511902581.004 Published online by Cambridge University Press