Monte Carlo Transition Dynamics and Variance Reduction

M. Fitzgerald,[1] **R. R. Picard,**[1] **and R. N. Silver**[1]

For Metropolis Monte Carlo simulations in statistical physics, efficient, easy-to-implement, and unbiased statistical estimators of thermodynamic properties are based on the transition dynamics. Using an Ising model example, we demonstrate (problem-specific) variance reductions compared to conventional histogram estimators. A proof of variance reduction in a microstate limit is presented.

1. BACKGROUND

Monte Carlo computer simulation has proven extremely useful for problems in statistical physics and chemistry. As originally suggested[1] by Metropolis *et al.*, a Markov chain of system microstates can be simulated whose limiting relative frequencies are equal to their Boltzmann probabilities. Thus, thermodynamic properties of the system can be estimated by simple averaging; that is, by collecting histograms. The use of histogram estimators has remained prevalent despite the many improvements to the original Metropolis algorithm over the years, such as non-Boltzmann sampling methods for slowly relaxing systems.

It may be surprising to many Monte Carlo practitioners to learn that there are more efficient statistical estimators for physical properties than those obtained from histograms. In this paper, we examine the statistical properties of new estimators[2] based on the transition dynamics of the system, using *canonical transition probabilities* (CTPs) estimated from the

---

[1] Los Alamos National Laboratory, Los Alamos, New Mexico 87545; e-mail: markfitz@ lanl.gov; picard@lanl.gov; rns@loke.lanl.gov.

Monte Carlo simulation. The prospect of variance reduction arises because the estimators of the transition dynamics use microstate-level information that has smaller variance than the microstate-level information used in histogram estimators.

The CTP method is easy to implement, amounting to the insertion of a simple bookkeeping step into the Metropolis algorithm. This bookkeeping is compatible with any Metropolis proposal mechanism, any definition of macrostate labels, and any importance weights guiding the simulation. The use of CTPs is easily extended[2] to other scenarios, such as those involving partitioning of the state space and/or adaptive updating of importance weights. Extrapolation of system properties to temperatures other than that of the simulation is also possible. Because the bookkeeping can simply be accumulated over adaptive runs, the adaptive capabilities of CTPs are appealing, especially for slowly relaxing systems, where it is crucial to adapt to good importance weights. Even in static simulations using Boltzmann importance weights, however, CTPs provide reduced variance, as we show in this paper.

The extent of variance reduction is problem-specific, depending on such issues as the thermodynamic quantity of interest, the system size, the temperature, and the implementation details for the Metropolis algorithm (e.g., the mechanism used to generate proposed moves and the importance weights used to guide the random walk). We present an Ising model example where the histogram estimators have variance roughly 1.25 times that achieved by CTP estimators of transition dynamics, i.e., it would require running the simulation 25% longer to obtain the same precision using histogram estimators. While such gains are not spectacular, they are applicable to most Metropolis Monte Carlo simulations in statistical physics and chemistry, and are available at essentially no cost.

An outline of this paper is as follows. First, we review the Metropolis algorithm and standard histogram estimators. Second, we discuss alternative estimators based on histograms. Third, we discuss estimators based on macrostate transition dynamics, which should reduce variance relative to either of the histogram estimators. (This argument uses a proof of variance reduction in the microstate limit, details of which are presented in an appendix and generalized to include the use of arbitrary importance weights guiding the random walk.) Finally, we empirically demonstrate variance reduction in an Ising model example.

2. STANDARD ESTIMATES

To establish notation, consider a system with a set Ω of possible states, where microstate $s \in \Omega$ has Boltzmann probability

$$\pi_s = \frac{e^{-\beta E_s}}{\sum_{t \in \Omega} e^{-\beta E_t}} = \frac{e^{-\beta E_s}}{Z} \tag{1}$$

Here, β denotes the inverse temperature of the system, E_s the energy associated with microstate s, and Z the partition function. In the Ising model, for example, a microstate s is a configuration of $\pm$ spins, $\{\sigma_a\}$, over a set of lattice points, and the energy E_s is the sum of nearest neighbor interactions, i.e., $E_s = -\sum_{(a,b) \in n.n.} \sigma_a \sigma_b$.

Most applications involve only the Boltzmann probabilities of macrostates, from which thermodynamic properties can be derived. Let S denote the macrostate corresponding to a set of microstates s sharing some characteristic of interest. The probability Π_S for macrostate S is

$$\Pi_S = \sum_{s \in S} \pi_s \tag{2}$$

Henceforth, we use lower-case letters to denote microstates and upper-case letters for macrostates. For example, in the Ising model, the macrostate of interest could denote magnetization, where microstate s is in macrostate S if its magnetization, $M_s = \sum_a \sigma_a^{(s)}$, equals M_S. A thermodynamic quantity of interest derived from this macrostate is the magnetic susceptibility

$$\chi \propto \sum_S \Pi_S M_S^2 \tag{3}$$

The Metropolis algorithm produces a Markov chain which, upon equilibration, visits microstates with relative sampling frequencies π_s and macrostates with relative frequencies Π_S. Each step of the Metropolis algorithm consists of three parts:

$\boxed{1}$ Given that the system is in microstate $s \in S$, a move to another microstate $t \in T$ is proposed by some mechanism. We denote

$$q_{s,t} = \text{Prob}(\text{state } t \text{ is proposed} \mid \text{system is in state } s). \tag{4}$$

To simplify the presentation, symmetry $q_{s,t} = q_{t,s}$ is assumed here, though this is not strictly necessary.[3]

$\boxed{2}$ Given that the system is in microstate $s \in S$ and a move to microstate $t \in T$ has been proposed, the probability that the system moves to state t is

$$r_{s,t} = \text{Prob}(\text{move to state } t \mid \text{system is in state } s \text{ and state } t \text{ is proposed})$$
$$= \min\{1, \pi_t/\pi_s\}. \tag{5}$$

If the system does not move to the proposed state t, then it remains in state s.

The third part of each Metropolis step consists of the bookkeeping used to estimate system properties of interest.

The commonly used histogram estimator counts the number of times that the chain visits each macrostate S. That is, upon equilibration of the Markov chain, the Metropolis algorithm invokes the bookkeeping step

> 3a   If the system moved to state $t \in T$, increment the histogram count $H_T = H_T + 1$; if the system remained in state $s \in S$, increment $H_S = H_S + 1$.

The histogram estimator of the macrostate probability Π_S is

$$\hat{\Pi}_S^{(H)} = \frac{H_S}{N} \tag{6}$$

where N denotes the number of equilibrated Metropolis steps in the simulation. **Note**: *Throughout this paper, the symbol "^" refers to a Monte Carlo estimate, which is to be contrasted with the actual quantity being estimated.* For our Ising example, this histogram estimate could be substituted into Eq. (3) to estimate the susceptibility. The average magnetization is obtained similarly. This procedure is entirely equivalent to averaging the time series of observed magnetizations M_S. Correlation properties of the stationary time series would then be used for error propagation.[4]

An alternative, less well-known, histogram estimator[5] for the Metropolis algorithm uses the *empirical transition probabilities* (ETPs) between macrostates and detailed balance. The corresponding bookkeeping step creates a histogram of the transitions that the system makes from one macrostate to another:

> 3b   If the system moved to state $t \in T$, increment the count $H_{S,T} = H_{S,T} + 1$; if the system remained in state $s \in S$, increment $H_{S,S} = H_{S,S} + 1$.

The bookkeeping for ETPs could, in principle, involve a large two-dimensional array. For many problems, however, much simplification occurs. In the Ising model with a single-spin-flip proposal mechanism, for example, the $\{H_{S,T}\}$ array is banded with three bands (if magnetization defines macrostates) or five bands (if energy defines macrostates). Even in cases where the full $\{H_{S,T}\}$ array can be populated, the bookkeeping can be simplified by accumulating only the counts for macrostate pairs $\{S, T\}$ in the primary bands, relegating the low counts $\{H_{S,T}\}$ from other bands into the diagonal $\{H_{S,S}\}$ terms. This approach is tantamount to using only the detailed balance equations from the primary bands in the estimation of macrostate probabilities. The reduced storage from using only the primary bands trades off with a loss of information, but this loss is slight when the counts in the other bands are small.

Note that the $\{H_{S,T}\}$ transition histograms contains all of the information from the standard histogram method, because

$$H_S = \sum_T H_{S,T}$$

Equilibrated transition probabilities $\{P_{S,T}\}$ for the Markov chain are estimated using ETPs by

$$\hat{P}_{S,T}^{(ETP)} = \frac{H_{S,T}}{\sum_U H_{S,U}} = \frac{H_{S,T}}{H_S} \tag{7}$$

ETP-based system properties can be derived by using the estimates $\{\hat{P}_{S,T}^{(ETP)}\}$ together with detailed balance,

$$\Pi_S P_{S,T} = \Pi_T P_{T,S} \tag{8}$$

This derivation involves taking logarithms in Eq. (8) to improve numerical stability, substituting the estimate $\hat{P}_{S,T}^{(ETP)}$ for $P_{S,T}$, rearranging terms, and producing a set of equations in the logged ETP-based estimates of the $\{\Pi_S\}$:

$$\ln \hat{P}_{T,S}^{(ETP)} - \ln \hat{P}_{S,T}^{(ETP)} = \ln \hat{\Pi}_S^{(ETP)} - \ln \hat{\Pi}_T^{(ETP)} \tag{9}$$

for all (S,T) for which the detailed balance, Eq. (8), is not of the form $0 = 0$ (for some pairs (S,T), $P_{S,T} = 0$ because the proposal mechanism does not allow such one-step transitions). This set of linear equations can be solved—a simple approach being to use standard software routines[6] for least squares to minimize

$$\sum_{S,T} ([\ln \hat{P}_{T,S}^{(ETP)} - \ln \hat{P}_{S,T}^{(ETP)}] - [\ln \hat{\Pi}_S^{(ETP)} - \ln \hat{\Pi}_T^{(ETP)}])^2$$

and get ETP-based estimates $\hat{\Pi}_S^{(ETP)}$.

Any variance reduction through the use of ETPs relative to the standard histogram approach arises from more efficient use of the "data" $\{H_{S,T}\}$ from the simulation. As shown in the appendix, however, when the observed $\{H_{S,T}\}$ approximately satisfy an empirical balance condition $H_{S,T} = H_{T,S}$ for all $S \neq T$, then the variance reduction relative to the standard histogram method is minimal. In certain problems, such as the Ising model with magnetization defining the macrostates and single-spin-flip proposal mechanisms, approximate empirical balance always holds and ETP-based estimates are virtually identical to those for the standard histogram method.

3. TRANSITION DYNAMICS BASED ON ACTUAL TRANSITION PROBABILITIES

Improvement on histogram methods is possible with the CTP approach[2] that we now describe. This approach invokes the bookkeeping step

> $\boxed{\text{3c}}$ Whether or not the system moved to state $t \in T$, increment $C_{S,T} = C_{S,T} + r_{s,t}$ as well as $C_{S,S} = C_{S,S} + [1 - r_{s,t}]$; (note: if $S = T$, this means that the updating is $C_{S,S} = C_{S,S} + 1$).

The term $r_{s,t}$, defined in Eq. (5), is the actual transition probability of a move from state s to state t given that such a move has been proposed. This bookkeeping step $\boxed{\text{3c}}$ increments an array $C_{S,T}$ by $r_{s,t}$ and $1 - r_{s,t}$, in contrast to histogram methods that increment by zeroes and ones. As may be intuitively obvious, the use of actual probabilities can be a significant source of variance reduction.

The corresponding estimate of the canonical transition probability, $P_{S,T}$, arising from the $C_{S,T}$ array is similar to Eq. (7),

$$\hat{P}_{S,T}^{(CTP)} = \frac{C_{S,T}}{\sum_U C_{S,U}} = \frac{C_{S,T}}{H_S} \tag{10}$$

The Boltzmann macrostate probabilities are obtained from the estimated transition probabilities through solution of the detailed balance equations (8), just as for the ETP method.

In order to compare the relative efficiency of the CTP method to the histogram approaches, consider the microstate-level contributors $h_{s,t}$ (the number of transitions made from microstate s to microstate t) to the macrostate-level $H_{S,T}$ array

$$H_{S,T} = \sum_{s \in S} \sum_{t \in T} h_{s,t}$$

and the corresponding contributor $c_{s,t}$ to the $C_{S,T}$ array

$$C_{S,T} = \sum_{s \in S} \sum_{t \in T} c_{s,t}$$

As with denoting system states, the lower-case $h_{s,t}$ and $c_{s,t}$ is used to distinguish microstate-level quantities from macrostate-level quantities. The expected values

$$E[c_{s,t}] = E[h_{s,t}] = \pi_s q_{s,t} r_{s,t} \qquad \text{for all microstates } s \neq t$$

Loosely speaking, this says that $c_{s,t}$ and the corresponding histogram count $h_{s,t}$ of the observed transitions from s to t are estimating the same thing, but in different ways. The histogram $h_{s,t}$ counts the number of times that the system actually moved from s to t, while $c_{s,t}$ counts the number of times that a move from s to t was proposed and then multiplies by the acceptance probability for the proposal. The simulated Markov chain is the same in both cases, but the bookkeeping involved is different.

In the appendix, we prove that the direct use of acceptance probabilities $r_{s,t}$ leads to reduced variance:

$$\mathrm{Var}(c_{s,t}) \leqslant \mathrm{Var}(h_{s,t}) \qquad \text{for all microstates } s \neq t \qquad (11)$$

Equality holds in Eq. (11) if and only if $r_{s,t} = 1$ (then, proposed moves from s to t are always accepted, and $c_{s,t} \equiv h_{s,t}$) or when $q_{s,t} = 0$ (then, moves from s to t are never proposed, and $c_{s,t} \equiv h_{s,t} \equiv 0$).

The variance inequality, Eq. (11), implies that the building blocks of the CTP estimators are uniformly better than those of histogram methods. Nonetheless, this result falls somewhat short of showing that the macrostate density-of-states estimates based on the $\{C_{S,T}\}$ are superior to those based on $\{H_{S,T}\}$, which is necessary to complete the proof that the approach described herein is superior to histogram methods. Such a conjecture is plausible, however. Extending the variance inequality to the macrostate level is difficult because macrostate transitions do not have Markovian behavior, so that the proof given in the appendix does not apply.

There is a similarity between the CTP approach and two other approaches to the general problem: Transition Matrix Monte Carlo Reweighting[7] and Broad Histogram Monte Carlo.[8] In those approaches, two-dimensional arrays are also stored, which accumulate at each visited state the total numbers of single-spin-flip transitions to lower and higher energies. This information is then used to estimate the density of states, using ideas similar to those in converting CTPs. No theoretical variance reduction is shown for either of those approaches relative to the standard approach; indeed, any empirical reduction in variance per step of the random walk must be interpreted in light of the increased computational work required to assess all possible transitions from each visited state.

To illustrate the use of CTPs, we evaluated the variance reduction in a 30×30 Ising system slightly above the critical temperature, simulating for $\beta = 0.42$ and using the single-spin-flip proposal mechanism. This Ising system was chosen to be large enough to provide meaningful results, yet small enough to allow for the extensive simulation required to conclusively demonstrate variance reduction. Separate calculations were carried out for

magnetic susceptibility (using magnetization to define the macrostates) and specific heat (using energy to define macrostates). In each case, system properties were determined *from the same simulation data* using the various bookkeeping methods.

Figure 1 displays the results from 500 simulations for magnetic susceptibility, each simulation carried out for 5×10^6 sweeps of the system (a "sweep" of the 30×30 Ising system being equal to 900 Metropolis steps). Plotted is the average squared error as a function of iteration number, where each iteration denotes 10,000 sweeps. That is, a plotted point is the sample variance of the 500 simulated values. Superimposed on the results is a smooth curve, equal to a fitted constant times $1/N$, which can be used to quantify variance reduction. Upon comparing the curves for CTPs and for the standard histogram method (as is shown in Appendix A.2, the standard histogram method is essentially identical to the ETP approach in this case), the variance difference is 25 %. In other words, determining magnetic susceptibility by the standard approach would require running the simulation 1.25 times as long in order to achieve the same statistical accuracy.
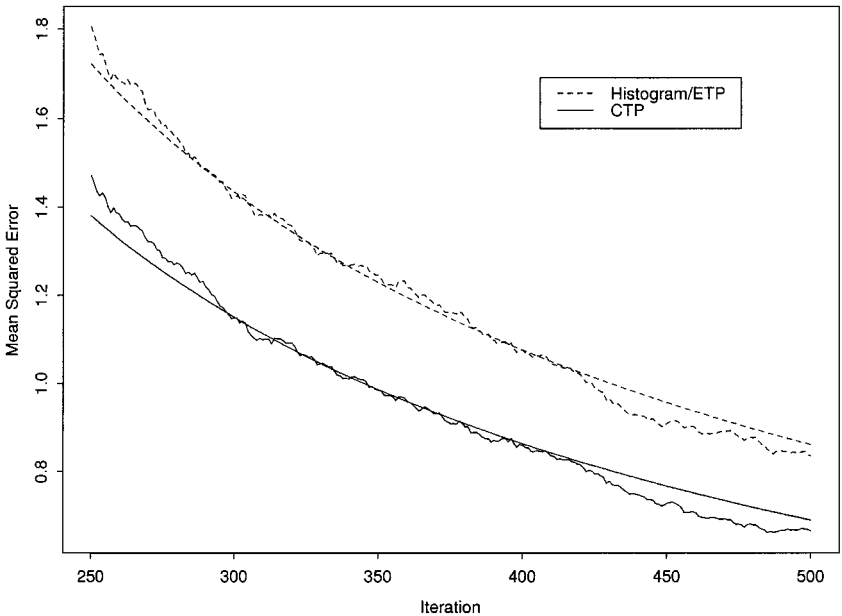

Fig. 1. Comparison of Mean Squared Error for histogram and CTP estimates of magnetic susceptibility in a 30×30 Ising model simulated at $\beta = 0.42$. Relative efficiency of CTP estimator is 1.25.

Another set of 500 simulations was carried out using energy to define system macrostates, comparing the approaches in their determination of the system's specific heat. Figure 2 displays the squared errors between the simulated estimates and the known specific heat.[9] The variance reduction from CTPs in this case is modest, roughly 6–8 % as compared with the standard histogram and ETP methods. This reduction is considerably less than that for magnetic susceptibility, illustrating the problem-specific nature of the phenomenon.

In evaluating variance reductions, computational effort must be considered. Implementation of CTPs involves additional storage relative to the standard histogram approach ($C_{S,T}$ being a two-dimensional array) as well as the solution of the detailed balance equations, Eq. (9). For many proposal mechanisms $q_{s,t}$, however, the two-dimensional array reduces to a band system, so that storage needs are modest. Moreover, the balance equations must be solved only once (at the end of a simulation), and efficient software exists for this purpose. In our simulations, for example, a CPU profiling indicated that less than 0.1 % of the overall runtime was
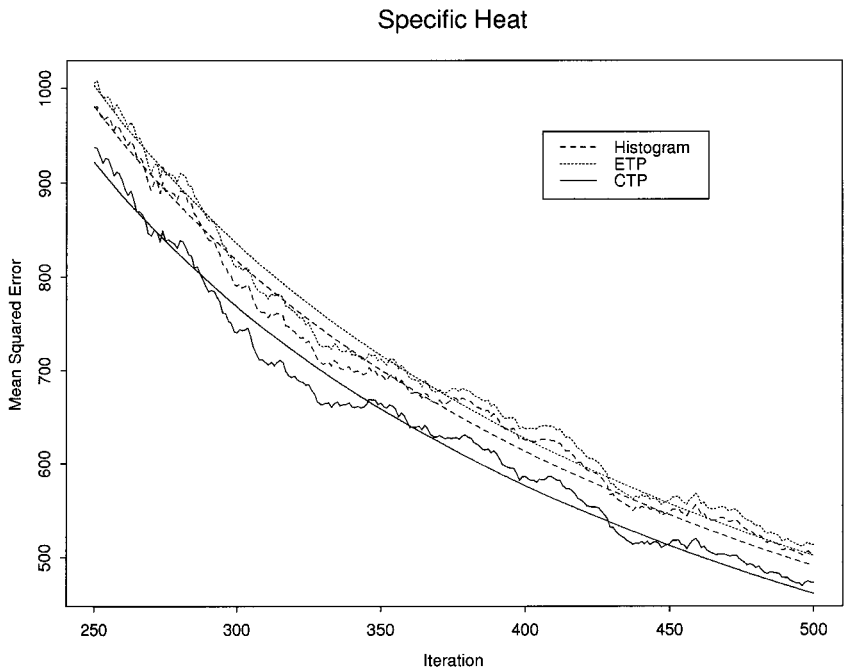
Specific Heat

Fig. 2. Comparison of Mean Squared Error for histogram, ETP, and CTP estimates of specific heat in a 30×30 Ising model simulated at $\beta = 0.42$. Relative efficiency of CTP estimator is 1.06 and 1.08 versus the histogram and ETP methods, respectively.

devoted to solving the detailed balance equations and producing estimates of system properties.

4. CONCLUSION

We have described improved statistical estimators based on Monte Carlo transition dynamics, which are widely applicable to diverse simulations in statistical physics and chemistry. While we have not (yet) proven theoretical variance reduction in the most general case, we have shown that the microstate building blocks for CTP estimators have reduced variance compared with the building blocks of the standard histogram estimators; moreover, this is true for non-Boltzmann importance-weighted simulations as well as for Boltzmann sampling (see the appendix). We have also demonstrated variance reduction empirically with an Ising model example.

That variance is reduced through use of CTPs offers great potential for adaptive Monte Carlo. Consider the case where multiple iterations of an adaptive algorithm learn desirable importance weights. With each iteration, variance is reduced as per the results of this paper. And because each iteration of the learning algorithm builds on its predecessors, variance reductions accrue in the vein of compound interest. Moreover, when energy is used to define the macrostates, either alone or in conjunction with another variable, Boltzmann transition probabilities can be simply accumulated over all iterations of the adaptive algorithm and used to estimate the density of states (see Appendix A.3).

The bookkeeping required by CTPs is simple and easy to implement. This bookkeeping is compatible with any Metropolis proposal mechanism, any definition of macrostate labels, and any importance weights guiding the simulation. Moreover, extrapolation of system properties to temperatures other than that of the simulation is also possible. These features, in addition to the reduced variance provided, make CTPs a useful tool in Monte Carlo simulation.

APPENDIX

A.1. Importance Sampling

For generality, the variance reduction proof is carried out for all importance-weighted Metropolis simulations, treating Boltzmann sampling as a special case. In order to do this, a brief review of importance sampling is needed.

The Boltzmann probability for microstate $s \in \Omega$ is now denoted (compare with Eq. (1))

$$\pi_s(\vec{0}) = \frac{e^{-\beta E_s}}{\sum_{t \in \Omega} e^{-\beta E_t}} = \frac{e^{-\beta E_s}}{Z}$$

and that for macrostate S is (compare with Eq. (2))

$$\Pi_S(\vec{0}) = \sum_{s \in S} \pi_s(\vec{0})$$

where the notation $\vec{0}$ is explained shortly. An importance-weighted Metropolis simulation produces a Markov chain which, upon equilibration, visits microstates with relative sampling frequencies $\pi_s(\vec{\eta}) \propto e^{\eta_S} \pi_s(\vec{0})$ and macrostates with relative frequencies $\Pi_S(\vec{\eta}) \propto e^{\eta_S} \Pi_S(\vec{0})$. Here, the vector of weights $\vec{\eta}$ defining the importance sampling depends only on the macrostate labels, i.e., $\eta_s \equiv \eta_S$ for all $s \in S$. For example, the choice $\vec{\eta} = \vec{0}$ corresponds to Boltzmann importance weights, while the choice $\eta_S = -\ln \Pi_S(\vec{0})$ corresponds to a multicanonical approach,[10] where the sampling frequencies for all macrostates are equal.

The acceptance portion of the Metropolis algorithm for importance-weighted random walks is

2′   Given that the system is in microstate $s \in S$ and a move to microstate $t \in T$ has been proposed, the probability that the system moves to state t is (compare with Eq. (5)):

$$r_{s,t}(\vec{\eta}) = \text{Prob(move to state } t \mid \text{system is in state } s \text{ and state } t \text{ is proposed)}$$

$$= \min\{1, e^{\eta_T} \pi_t(\vec{0})/e^{\eta_S} \pi_s(\vec{0})\}$$

$$= \min\{1, e^{-\beta(E_t - E_s) + (\eta_T - \eta_S)}\}$$

If the system does not move to the proposed state t, then it remains in state s.

For the standard histogram method, the number of times H_S that the chain is in S is proportional to the macrostate probability $\Pi_S(\vec{\eta}) \propto e^{\eta_S} \Pi_S(\vec{0})$; that is, histogram estimates are (compare with Eq. (6))

$$\hat{\Pi}_S^{(H)}(\vec{\eta}) = \frac{H_S}{N} \tag{12}$$

Histogram estimates of Boltzmann probabilities are obtained by inverting the importance sampling relation $\Pi_S(\vec{\eta}) \propto e^{\eta_S}\Pi_S(\vec{0})$ to give

$$\hat{\Pi}_S^{(H)}(\vec{0}) = \frac{e^{-\eta_S}\hat{\Pi}_S^{(H)}(\vec{\eta})}{\sum_T e^{-\eta_T}\hat{\Pi}_T^{(H)}(\vec{\eta})} = \frac{e^{-\eta_S}H_S}{\sum_T e^{-\eta_T}H_T} \tag{13}$$

For the ETP approach, the solution for equilibrated transition probabilities $\{P_{S,T}(\vec{\eta})\}$ for the importance-weighted Markov chain are estimated using ETP's by (compare with Eq. (7)):

$$\hat{P}_{S,T}^{(ETP)}(\vec{\eta}) = \frac{H_{S,T}}{\sum_U H_{S,U}} = \frac{H_{S,T}}{H_S} \tag{14}$$

ETP-based system properties can be derived by using the estimates $\{\hat{P}_{S,T}^{(ETP)}(\vec{\eta})\}$ together with the detailed balance, (compare with Eq. (8))

$$\Pi_S(\vec{\eta})\, P_{S,T}(\vec{\eta}) = \Pi_T(\vec{\eta})\, P_{T,S}(\vec{\eta}) \tag{15}$$

This derivation involves solving a set of equations in the logged ETP-based estimates of the $\{\Pi_S(\vec{\eta})\}$ (compare with Eq. (9)):

$$\ln \hat{P}_{T,S}^{(ETP)}(\vec{\eta}) - \ln \hat{P}_{S,T}^{(ETP)}(\vec{\eta}) = \ln \hat{\Pi}_S^{(ETP)}(\vec{\eta}) - \ln \hat{\Pi}_T^{(ETP)}(\vec{\eta}) \tag{16}$$

for all (S, T) for which the detailed balance, Eq. (15), is not of the form $0 = 0$. This set of linear equations can be solved as in the Boltzmann case to get ETP-based estimates $\hat{\Pi}_S^{(ETP)}(\vec{\eta})$. The estimated macrostate probabilities are converted to estimated Boltzmann probabilities $\hat{\Pi}_S^{(ETP)}(\vec{0})$ as in Eq. (13).

Under importance sampling, transition dynamics correspond to the bookkeeping step

> 3c   Whether or not the system moved to state $t \in T$, increment $C'_{S,T} = C'_{S,T} + r_{s,t}(\vec{\eta})$ as well as $C'_{S,S} = C'_{S,S} + [1 - r_{s,t}(\vec{\eta})]$ (note: if $S = T$, this means that the updating is $C'_{S,S} = C'_{S,S} + 1$).

The term $r_{s,t}(\vec{\eta})$ in this bookkeeping step is defined in Metropolis stage 2′ and is the actual transition probability of a move from state s to state t given that such a move has been proposed. Further, the notation C' distinguishes $C'_{S,T}$ from the $C_{S,T}$ in 3c referenced to the canonical weights $\vec{\eta} = \vec{0}$.

A.2. ETP and Standard Histogram Estimates

In this section, we show that when the counts $H_{S,T}$ satisfy an empirical detailed balance condition that ETP estimates are identical to those

from the usual histogram method. A corollary of this result is that if empirical detailed balance is nearly satisfied, then the ETP and standard histogram methods give nearly identical results.

Recall that the detailed balance condition is given by Eq. (15),

$$\Pi_S(\vec{\eta}) \, P_{S,\,T}(\vec{\eta}) = \Pi_T(\vec{\eta}) \, P_{T,\,S}(\vec{\eta}) \qquad \text{for all macrostate pairs } (S, T)$$

Viewed somewhat differently, this is a statement about the expected values of the transition counts $H_{S,\,T}$ of the Markov chain:

$$E(H_{S,\,T}) = E(H_{T,\,S}) \qquad \text{for all macrostate pairs } (S, T)$$

Suppose that the empirical version of this were to hold in a particular simulation of N equilibrated steps, i.e.,

$$H_{S,\,T} = H_{T,\,S} \qquad \text{for all macrostate pairs } (S, T) \tag{17}$$

Admittedly, the probability of all counts $\{H_{S,\,T}\}$ satisfying these equalities exactly is very small for many proposal mechanisms, but the purpose here is to examine the relation of ETP estimates to those from the standard histogram approach.

Using empirical detailed balance,

$$
\begin{aligned}
\ln \left(\frac{\hat{P}_{T,\,S}^{(ETP)}(\vec{\eta})}{\hat{P}_{S,\,T}^{(ETP)}(\vec{\eta})} \right) &= \ln \left(\frac{H_{T,\,S}/H_T}{H_{S,\,T}/H_S} \right) \\
&= \ln(H_S/H_T) \\
&= \ln H_S - \ln H_T
\end{aligned}
\tag{18}
$$

Substituting Eq. (18) into the set of linear equations (16) to be solved then gives

$$\ln \left(\frac{\hat{P}_{T,\,S}^{(ETP)}(\vec{\eta})}{\hat{P}_{S,\,T}^{(ETP)}(\vec{\eta})} \right) = \ln H_S - \ln H_T = \ln \hat{\Pi}_S^{(ETP)}(\vec{\eta}) - \ln \hat{\Pi}_T^{(ETP)}(\vec{\eta}) \tag{19}$$

Thus, when empirical detailed balance holds for all macrostate pairs (S, T), the (exact) solution to the linear equations (19) is easily seen to be

$$\hat{\Pi}_S^{(ETP)}(\vec{\eta}) = \frac{H_S}{N}$$

which is the same set of estimates as obtained from the standard histogram method, Eq. (12).

In other words, the difference between ETP estimates and those from the usual histogram approach reflects the extent to which empirical detailed balance in Eq. (17) is not satisfied.

For a certain class of problems, approximate empirical detailed balance holds, almost by definition. As an extreme example, consider the Ising model with magnetization defining the macrostates. When the proposal mechanism $q_{s,t}$ is equivalent to selecting a single lattice point and flipping its spin, the system moves at most one macrostate per Metropolis step. Note that this implies that for all adjacent macrostate pairs (S, T), either $H_{S,T} = H_{T,S}$ or $H_{S,T} = H_{T,S} \pm 1$. It is impossible for $H_{S,T}$ to be more than 1 from $H_{T,S}$ because any time that the random walk moves from S to T, it can only return again to state S via state T—recall that the system can move no more than one state per step in this case.

Because the empirical detailed balance condition is "almost" satisfied for this problem (to within ± 1 in a Markov chain whose length N is usually many thousands of steps), Eq. (17) "almost" holds and ETP-based estimates are virtually identical to those for the standard histogram method. Any variance reduction through use of ETP's is obviously minimal here, and not worth the computational overhead of maintaining an $H_{S,T}$ array.

A.3. Bookkeeping in Canonical Scale

We now show that when the system Hamiltonian is used to define the macrostates, either alone or in conjunction with another relevant variable, that carrying out the bookkeeping 3c with the canonical acceptance probabilities $r_{s,t}(\vec{0})$ gives results identical to those for the bookkeeping 3c' using acceptance probabilities $r_{s,t}(\vec{\eta})$ referenced to the weights $\vec{\eta}$ guiding the simulation. Thus, canonical estimates are identical to those based on the transition dynamics $\{C'_{S,T}\}$ of the simulation. This simplification allows for accumulating the canonical array $C_{S,T}$ over several simulation runs having (possibly) different weights $\vec{\eta}$, thereby taking advantage of parallel computing and making the approach compatible with adaptive methods, while preserving reduced variance.

The set of equations to be solved in an importance-weighted simulation, Eq. (16), can be written in matrix form as

$$\vec{y}_{\vec{\eta}} = \mathbf{X}\vec{\theta}_{\vec{\eta}} \tag{20}$$

where the components of $\vec{y}_{\vec{\eta}}$ have the generic form of logged ratios of transition probability estimates (see Eqs. (9) and (10)),

$$\ln(\hat{P}^{(CTP')}_{T,S}(\vec{\eta})/\hat{P}^{(CTP')}_{S,T}(\vec{\eta}))$$

The components of $\vec{\theta}_{\vec{\eta}}$ are the unnormalized logged macrostate probabilities $\ln \Pi_S(\vec{\eta})$, and the design matrix $\mathbf{X}$ has rows containing a single $+1$, a single -1, and all other entries equal to zero.

Now consider the canonical bookkeeping step $\boxed{3c}$ described earlier:

$$C_{S,T} = C_{S,T} + r_{s,t}(\vec{0}) \qquad \text{and} \qquad C_{S,S} = C_{S,S} + [1 - r_{s,t}(\vec{0})] \qquad (21)$$

At the end of the importance-weighted simulation, let $n_{S,T}$ denote the number of Metropolis steps with the system in macrostate S with macrostate T proposed. The canonical counting array satisfies

$$C_{S,T} = n_{S,T} r_{s,t}(\vec{0}) \qquad (22)$$

where $r_{s,t}(\vec{0}) = \min\{1, \pi_t(\vec{0})/\pi_s(\vec{0})\}$ is the acceptance probability for the proposed move were Boltzmann sampling to be used and were the system in state s with state t proposed. When energy is used to define the macrostates, the acceptance probability $r_{s,t}(\vec{0})$ is the same for all $s \in S$ and $t \in T$, allowing for the simplification in Eq. (22).

Note that the Boltzmann acceptance probabilities satisfy the relation

$$\begin{aligned}
\frac{r_{t,s}(\vec{0})}{r_{s,t}(\vec{0})} &= \frac{\min\{1, \pi_s(\vec{0})/\pi_t(\vec{0})\}}{\min\{1, \pi_t(\vec{0})/\pi_s(\vec{0})\}} \\
&= \frac{\pi_s(\vec{0})}{\pi_t(\vec{0})} \\
&= \left(\frac{e^{\eta_T}}{e^{\eta_S}}\right)\left(\frac{e^{\eta_S}\pi_s(\vec{0})}{e^{\eta_T}\pi_t(\vec{0})}\right) \\
&= \left(\frac{e^{\eta_T}}{e^{\eta_S}}\right)\frac{\min\{1, \pi_s(\vec{\eta})/\pi_t(\vec{\eta})\}}{\min\{1, \pi_t(\vec{\eta})/\pi_s(\vec{\eta})\}} \\
&= \left(\frac{e^{\eta_T}}{e^{\eta_S}}\right)\left(\frac{r_{t,s}(\vec{\eta})}{r_{s,t}(\vec{\eta})}\right) \qquad (23)
\end{aligned}$$

Moreover, $\sum_T C_{S,T} = \sum_T C'_{S,T} = H_S$, which implies

$$\hat{P}^{(CTP)}_{S,T}(\vec{0}) = n_{S,T} r_{s,t}(\vec{0})/H_S$$

Using this relation and Eq. (23),

$$\ln\left(\frac{\hat{P}_{T,S}^{(CTP)}(\vec{0})}{\hat{P}_{S,T}^{(CTP)}(\vec{0})}\right) = \ln\left(\frac{n_{T,S}\,r_{t,s}(\vec{0})/H_T}{n_{S,T}\,r_{s,t}(\vec{0})/H_S}\right)$$

$$= \ln\left(\frac{n_{T,S}/H_T}{n_{S,T}/H_S}\right) + \ln\left(\frac{r_{t,s}(\vec{0})}{r_{s,t}(\vec{0})}\right)$$

$$= \ln\left(\frac{n_{T,S}/H_T}{n_{S,T}/H_S}\right) + \ln\left[\left(\frac{e^{\eta_T}}{e^{\eta_S}}\right)\left(\frac{r_{t,s}(\vec{\eta})}{r_{s,t}(\vec{\eta})}\right)\right]$$

$$= (\eta_T - \eta_S) + \ln\left(\frac{n_{T,S}\,r_{t,s}(\vec{\eta})/H_T}{n_{S,T}\,r_{s,t}(\vec{\eta})/H_S}\right)$$

$$= (\eta_T - \eta_S) + \ln\left(\frac{\hat{P}_{T,S}^{(CTP')}(\vec{\eta})}{\hat{P}_{S,T}^{(CTP')}(\vec{\eta})}\right)$$

Comparing with Eq. (20), the components of $\vec{y}_{\vec{0}}$ are the shifted (by $\eta_T - \eta_S$) counterparts of those of $\vec{y}_{\vec{\eta}}$. It follows that the solution of the canonical set of linear equations $\vec{y}_{\vec{0}} = \mathbf{X}\vec{\theta}_{\vec{0}}$ is the same as that for $\vec{y}_{\vec{\eta}} = \mathbf{X}\vec{\theta}_{\vec{\eta}}$, except that the estimated log macrostate probabilities are shifted in the same way, i.e.,

$$\ln\hat{\Pi}_S^{(CTP)}(\vec{0}) - \ln\hat{\Pi}_T^{(CTP)}(\vec{0}) = (\eta_T - \eta_S) + (\ln\hat{\Pi}_S^{(CTP')}(\vec{\eta}) - \ln\hat{Pi}_T^{(CTP')}(\vec{\eta}))$$

which is equivalent to $\hat{\Pi}_S^{(CTP)}(\vec{0}) \propto e^{-\eta_S}\hat{\Pi}_S^{(CTP')}(\vec{\eta})$.

In other words, doing the bookkeeping in canonical scale via bookkeeping step ⌐3c⌐ for an importance-weighted simulation produces exactly the same estimates as doing the bookkeeping in $\vec{\eta}$-scale via ⌐3c'⌐ and then converting to canonical scale by Eq. (13). Note that the use of energy in defining the macrostates leads to Eq. (22), which is crucial in proving this result.

Finally, it is noted that the approach to solving the system of equations used in the Ising example, ordinary least squares, is easy to implement but is by no means optimal. Weighted least squares would give better variance reduction, if the proper weights could be determined.

A.4. Primary Steps in the Variance Reduction Proof

Variance reduction at the microstate-level, Eq. (11), follows from Markov chain theory.[11] The major steps of the proof are given here, with more lengthy mathematical details relegated to later sections in this appendix for those interested.

Consider a fixed-length, importance-weighted simulation of a Markov chain in equilibrium (i.e., simulating a prescribed number N of equilibrated steps of the Metropolis algorithm). The limiting probability of microstate s is $\pi_s(\vec{\eta})$; indeed, the Metropolis algorithm is explicitly designed to provide this result. The *recurrence time* of state s is defined as the number of Metropolis steps between consecutive visits to state s. It follows from Markov chain theory that the mean μ_s of the recurrence time distribution is simply the reciprocal of the limiting probability, i.e., $\mu_s = 1/\pi_s(\vec{\eta})$.

When the length N of the equilibrated Markov chain is large, the number of visits h_s to state s is approximately normally distributed with mean

$$E(h_s) \sim N/\mu_s$$

and variance

$$\text{Var}(h_s) \sim N\sigma_s^2/\mu_s^3$$

where σ_s^2 denotes the variance of the recurrence time for state s and the notation "$\sim$" means that the ratio of the two sides converges to 1 as $N \to \infty$. For the simulated Metropolis-based Markov chain, the above reduces to $E(h_s) \sim N\pi_s(\vec{\eta})$ and $\text{Var}(h_s) \sim N\sigma_s^2\pi_s^3(\vec{\eta})$.

These results are the basis for determining system properties by the standard histogram method, e.g., Eq. (12). To be sure, asymptotic theory can't be applied indiscriminately (there exist Markov chains for which recurrence times do not have finite expectation), but this is not an issue for the systems of interest in this paper.

To obtain the variance inequality $\text{Var}(c'_{s,t}) \leqslant \text{Var}(h_{s,t})$ as per Eq. (11), it is useful to consider the so-called expanded Markov chains for the transition and proposal processes, respectively. The (random) steps of the random walk lead to some number of occurrences when state s is visited and state t is proposed as the next move, that number being denoted as $n_{s,t}$. Similarly, there is some number of occurrences when state s is visited and the next move is made to state t, that number being denoted as $h_{s,t}$.

Each expanded Markov chain has states that are indexed by the ordered pair (s,t). The state (s,t) in the expanded chain for the transition process means that the system has just moved from state s to state t, while the state (s,t) in the expanded chain for the proposal process means that the system has just proposed a move from state s to state t. The first chain corresponds to the $n_{s,t}$ process at the heart of the method described here, while the second chain corresponds to the $h_{s,t}$ process at the heart of the ETP method. The transition and proposal chains are related in the obvious

way (it's not possible for the system to move from state s to state $t \neq s$ without such a move having been proposed), and this relationship between the two expanded chains leads to the desired variance inequality.

Relative to the bookkeeping for the Metropolis algorithm, $h_{s,t}$ is simply the number of visits that the expanded chain for the transition process makes to state (s, t), while $c'_{s,t} = n_{s,t} r_{s,t}(\vec{\eta})$, where $n_{s,t}$ is the number of visits that the expanded chain for the proposal process makes to state (s, t). From this characterization, large sample properties follow from Markov chain theory. The corresponding expected values are

$$E(h_{s,t}) \sim N\pi_s(\vec{\eta}) \, q_{s,t} r_{s,t}(\vec{\eta})$$

and

$$\begin{aligned} E(c'_{s,t}) &= E(n_{s,t}, r_{s,t}(\vec{\eta})) \\ &= r_{s,t}(\vec{\eta})[\, E(n_{s,t})\,] \\ &\sim r_{s,t}(\vec{\eta})[\, N\pi_s(\vec{\eta}) \, q_{s,t}\,] \\ &= N\pi_s(\vec{\eta}) \, q_{s,t} r_{s,t}(\vec{\eta}) \\ &\sim E(h_{s,t}) \end{aligned}$$

which was noted earlier for the Boltzmann case. In other words, $h_{s,t}$ and $c'_{s,t}$ are estimating the same quantity.

The variance reduction proof is based on a recurrence time argument. Choose a pair of microstates s and t where, for the results to be meaningful, it is assumed that s and t are such that

(a) $q_{s,t} > 0$ (otherwise, a one-step move from s to t can never be proposed; then, both $h_{s,t}$ and $n_{s,t}$ are always zero) and

(b) $r_{s,t}(\vec{\eta}) < 1$ (otherwise, the acceptance probability of a proposed move to t from s is equal to 1, and proposing a move to t from s is the same as moving to t from s; then, $h_{s,t} \equiv n_{s,t}$).

Let τ_h denote the number of Metropolis steps between successive (s, t) transitions and let τ_n denote the number of Metropolis steps between successive (s, t) proposals. For $E(\tau_h)$ and $\mathrm{Var}(\tau_h)$ denoting the mean and variance of the recurrence time τ_h, for example, the asymptotic variance of $h_{s,t}$ satisfies

$$\mathrm{Var}(h_{s,t}) = N \, \mathrm{Var}(\tau_h)/[\, E(\tau_h)\,]^3 \qquad (24)$$

The steps in the proof of Eq. (11), with an explanation to follow:

$$
\begin{aligned}
\mathrm{Var}(h_{s,t}) &= N\,\mathrm{Var}(\tau_h)/[E(\tau_h)]^3 \\
&= N\,\mathrm{Var}(\tau_h)\,\pi_s^3(\vec{\eta})\,q_{s,t}^3\,r_{s,t}^3(\vec{\eta}) \\
&= r_{s,t}^2(\vec{\eta})[\,N\,\mathrm{Var}(\tau_n)\,\pi_s^3(\vec{\eta})\,q_{s,t}^3\,][\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_h)/\mathrm{Var}(\tau_n)\,] \\
&= r_{s,t}^2(\vec{\eta})[\,\mathrm{Var}(n_{s,t})\,] \times [\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_h)/\mathrm{Var}(\tau_n)\,] \\
&= [\,\mathrm{Var}(n_{s,t}\,r_{s,t}(\vec{\eta}))\,] \times [\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_h)/\mathrm{Var}(\tau_n)\,] \\
&= [\,\mathrm{Var}(c'_{s,t})\,] \times [\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_h)/\mathrm{Var}(\tau_n)\,] \\
&= \mathrm{Var}(c'_{s,t}) \times r_{s,t}(\vec{\eta})\left[\,\mathrm{Var}(\tau_n \mid O=t) + \frac{1-r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})}\,\mathrm{Var}(\tau_n \mid O=s)\right. \\
&\qquad \left. + \frac{1-r_{s,t}(\vec{\eta})}{r_{s,t}^2(\vec{\eta})}\,\{E(\tau_n \mid O=s)\}^2\right]\Big/\mathrm{Var}(\tau_n) \qquad (25) \\
&= \mathrm{Var}(c'_{s,t}) \times \left[\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_n \mid O=t) + [\,1-r_{s,t}(\vec{\eta})\,]\,\mathrm{Var}(\tau_n \mid O=s)\right. \\
&\qquad \left. + \frac{1-r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})}\,\{E(\tau_n \mid O=s)\}^2\right]\Big/\mathrm{Var}(\tau_n) \\
&> \mathrm{Var}(c'_{s,t}) \times \left[\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_n \mid O=t) + [\,1-r_{s,t}(\vec{\eta})\,]\,\mathrm{Var}(\tau_n \mid O=s)\right. \\
&\qquad \left. + \frac{1-r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})}\,[\,r_{s,t}^2(\vec{\eta})\{E(\tau_n \mid O=t)-E(\tau_n \mid O=s)\}\,]^2\right]\Big/\mathrm{Var}(\tau_n) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (26) \\
&= \mathrm{Var}(c'_{s,t}) \times [\,r_{s,t}(\vec{\eta})\,\mathrm{Var}(\tau_n \mid O=t) + [\,1-r_{s,t}(\vec{\eta})\,]\,\mathrm{Var}(\tau_n \mid O=s) \\
&\qquad + r_{s,t}(\vec{\eta})[\,1-r_{s,t}(\vec{\eta})\,]\{E(\tau_n \mid O=t)-E(\tau_n \mid O=s)\}^2]/\mathrm{Var}(\tau_n) \\
&= \mathrm{Var}(c'_{s,t}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (27)
\end{aligned}
$$

The second equation comes from the fact that the mean recurrence time for a state in an irreducible chain is equal to the reciprocal of the limiting probability for the state (the limiting probability for the $h_{s,t}$ chain is equal to $\pi_s q_{s,t} r_{s,t}(\vec{\eta})$); the third through sixth equations are simply algebra to isolate $\mathrm{Var}(c'_{s,t})$ from the rest; the seventh, Eq. (25), expresses the variance of the recurrence time τ_h using a conditioning argument described later in this appendix, where the notation O denotes the outcome of the proposed move from s to t with acceptance probability $r_{s,t}(\vec{\eta}) < 1$ (with "$O=s$" denoting that the proposed move was rejected and "$O=t$" denoting that the proposed move was accepted); the eighth is merely algebra; the ninth,

Eq. (26), follows from an inequality derived later in this appendix; the tenth is more algebra, and the eleventh, Eq. (27), expresses the variance of the recurrence time τ_n using a conditioning argument described in the next section.

The variance inequality $\text{Var}(c'_{s,t}) \leqslant \text{Var}(h_{s,t})$ for all distinct microstates s and t implies that the building blocks of CTP estimators are uniformly better than those of histogram methods. Extending the variance inequality to the macrostate level is difficult because macrostate transitions do not have Markovian behavior and dealing with the covariance structures[12, 13] algebraically in this vein is difficult.

A.5. Derivation of Eq. (27)

For the $n_{s,t}$ chain, the variance $\text{Var}(\tau_n)$ of τ_n is determined through a conditioning relation common to the statistical literature. That is, the variance of one random variable can be written[14] as the mean of the conditional variance with respect to another random variable plus the variance of the conditional mean.

Applying the conditioning relation to the recurrence time τ_n, where the conditioning is based on the outcome O of the proposed transition (which is equal to t if the proposed move is accepted and equal to s if not):

$$
\begin{aligned}
\text{Var}(\tau_n) = & \, E_O \, \text{Var}(\tau_n \mid O) + \text{Var}_O \, E(\tau_n \mid O) \\
= & \, r_{s,t}(\vec{\eta}) \, \text{Var}(\tau_n \mid O = t) + [1 - r_{s,t}(\vec{\eta})] \, \text{Var}(\tau_n \mid O = s) \\
& + r_{s,t}(\vec{\eta}) [E(\tau_n) - E(\tau_n \mid O = t)]^2 \\
& + [1 - r_{s,t}(\vec{\eta})][E(\tau_n) - E(\tau_n \mid O = s)]^2 \\
= & \, r_{s,t}(\vec{\eta}) \, \text{Var}(\tau_n \mid O = t) + [1 - r_{s,t}(\vec{\eta})] \, \text{Var}(\tau_n \mid O = s) \\
& + r_{s,t}(\vec{\eta})[r_{s,t}(\vec{\eta}) \, E(\tau_n \mid O = t) \\
& + [1 - r_{s,t}(\vec{\eta})] \, E(\tau_n \mid O = s) - E(\tau_n \mid O = t)]^2 \\
& + [1 - r_{s,t}(\vec{\eta})][r_{s,t}(\vec{\eta}) \, E(\tau_n \mid O = t) \\
& + [1 - r_{s,t}(\vec{\eta})] \, E(\tau_n \mid O = s) - E(\tau_n \mid O = s)]^2 \\
= & \, r_{s,t}(\vec{\eta}) \, \text{Var}(\tau_n \mid O = t) + [1 - r_{s,t}(\vec{\eta})] \, \text{Var}(\tau_n \mid O = s) \\
& + r_{s,t}(\vec{\eta})[[1 - r_{s,t}(\vec{\eta})]^2 \, [E(\tau_n \mid O = s) - E(\tau_n \mid O = t)]]^2 \\
& + [1 - r_{s,t}(\vec{\eta})][r_{s,t}(\vec{\eta})]^2 \, [E(\tau_n \mid O = t) - E(\tau_n \mid O = s)]]^2 \\
= & \, r_{s,t}(\vec{\eta}) \, \text{Var}(\tau_n \mid O = t) + [1 - r_{s,t}(\vec{\eta})] \, \text{Var}(\tau_n \mid O = s) \\
& + r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})][E(\tau_n \mid O = t) - E(\tau_n \mid O = s)]^2
\end{aligned}
$$

which is Eq. (27) in the previous section. To explain the above, the first equation is simply the general variance decomposition; the second follows from the probability that $O = t$ equals $r_{s,t}(\vec{\eta})$ and the probability that $O = s$ is $[1 - r_{s,t}(\vec{\eta})]$; the rest is algebra to condense the expression.

A.6. Derivation of Eq. (25)

Consider the $h_{s,t}$ chain, and let $\tilde{O}$ denote the outcome that the first i proposed (s, t) moves are rejected and the $(i+1)$st proposed (s, t) move is accepted (where $i = 0, 1, ...$). Use of the conditioning on $\tilde{O}$ allows for coupling the recurrence time τ_h with the recurrence time τ_n for the $n_{s,t}$ chain. Note that the probability that $\tilde{O} = i$ follows a geometric distribution, and is equal to $r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i$.

Two important mathematical identities to be used here are

$$\sum_{i=1}^{\infty} i[1 - r_{s,t}(\vec{\eta})]^i = [1 - r_{s,t}(\vec{\eta})]/r_{s,t}^2(\vec{\eta})$$

and

$$\sum_{i=1}^{\infty} i^2[1 - r_{s,t}(\vec{\eta})]^i = [1 - r_{s,t}(\vec{\eta})][2 - r_{s,t}(\vec{\eta})]/r_{s,t}^3(\vec{\eta})$$

These identities follow from the moments of the geometric distribution; alternatively, they can be derived upon writing the expressions of interest as sums of derivatives of individual terms with respect to the index i, and then interchanging the order of summation and differentiation.

Using these identities and following the conditioning argument as in the previous section gives the variance of the recurrence time τ_h for the $h_{s,t}$ chain as

$$\begin{aligned}
\mathrm{Var}(\tau_h) &= E_{\tilde{O}}\,\mathrm{Var}(\tau_h \mid \tilde{O}) + \mathrm{Var}_{\tilde{O}}\,E(\tau_h \mid \tilde{O}) \\
&= \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i\,\mathrm{Var}(\tau_h \mid \tilde{O} = i) \\
&\quad + \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i\,[E(\tau_h \mid \tilde{O} = i) - E(\tau_h)]^2 \\
&= \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i\,[\mathrm{Var}(\tau_n \mid O = t) + i\,\mathrm{Var}(\tau_n \mid O = s)] \\
&\quad + \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i\,[E(\tau_n \mid O = t) + iE(\tau_n \mid O = s) - E(\tau_h)]^2
\end{aligned}$$

$$= \mathrm{Var}(\tau_n \mid O = t) + \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})} \mathrm{Var}(\tau_n \mid O = s)$$

$$+ \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i \Bigg[E(\tau_n \mid O = t) + iE(\tau_n \mid O = s)$$

$$- E(\tau_n \mid O = t) - \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})} E(\tau_n \mid O = s) \Bigg]^2$$

$$= \mathrm{Var}(\tau_n \mid O = t) + \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})} \mathrm{Var}(\tau_n \mid O = s)$$

$$+ [E(\tau_n \mid O = s)]^2 \sum_{i=0}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^i \left(i - \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})} \right)^2$$

$$= \mathrm{Var}(\tau_n \mid O = t) + \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}(\vec{\eta})} \mathrm{Var}(\tau_n \mid O = s)$$

$$+ \frac{1 - r_{s,t}(\vec{\eta})}{r_{s,t}^2(\vec{\eta})} [E(\tau_n \mid O = s)]^2$$

which is Eq. (25). The first equality above follows from the conditioning argument; the second from use of the geometric distribution for $\tilde{O}$; the third from the relation between the outcomes $\tilde{O}$ and O of the previous section; and rest from the aforementioned mathematical identities.

A.7. Derivation of Eq. (26)

To show Eq. (26), two results regarding passage times for simulated system are required. The first is that the expected passage time $E(s \to t$, no return to $s)$ from state s to state t, given that no intervening return to state s has occurred, is equal to the passage time $E(t \to s$, no return to $t)$. This result is a consequence of the detailed balance (sometimes called time reversibility), in that each path containing events $\{s \to t$, no return to $s\}$ has the same equilibrated probability as the reversed path containing events $\{t \to s$, no return to $t\}$.

The second result, perhaps less intuitive, is that the expectation $E(s \to t)$ of the number of Metropolis steps needed to move from state s to state t for the system is equal to $E(t \to s)$ of the number of steps needed to move from t to s. This follows from properties of first passage times. Letting $_t f_{ts}$ denote the probability of going from t to s without a return to t, we have[11]

$$_t f_{ts} = {}_s f_{st}(\pi_s / \pi_t)$$

Using this relation and the properties of the geometric distribution as per the previous section,

$$
\begin{aligned}
E(t \to s) &= \sum_{i=0}^{\infty} {}_t f_{ts} [1 - {}_t f_{ts}]^i \, [\, E(t \to s, \text{no return to } t) \\
&\quad + i E(t \to t, \text{no visit to } s)\,] \\[2mm]
&= E(t \to s, \text{no return to } t) + {}_t f_{ts} E(t \to t, \text{no visit to } s) \sum_{i=0}^{\infty} i [1 - {}_t f_{ts}]^i \\[2mm]
&= E(t \to s, \text{no return to } t) + {}_t f_{ts} E(t \to t, \text{no visit to } s) \frac{[1 - {}_t f_{ts}]}{[{}_t f_{ts}]^2} \\[2mm]
&= E(t \to s, \text{no return to } t) + \frac{[1 - {}_t f_{ts}]}{{}_t f_{ts}} E(t \to t, \text{no visit to } s) \\[2mm]
&= E(t \to s, \text{no return to } t) + \frac{[1 - {}_t f_{ts}]}{{}_t f_{ts}} \frac{1}{\pi_t [1 - {}_t f_{ts}]} \\[2mm]
&= E(t \to s, \text{no return to } t) + \frac{1}{\pi_t \, {}_t f_{ts}} \\[2mm]
&= E(s \to t, \text{no return to } s) + \frac{1}{\pi_t \, {}_s f_{st}(\pi_s / \pi_t)} \\[2mm]
&= E(s \to t, \text{no return to } s) + \frac{1}{\pi_s \, {}_s f_{st}} \\[2mm]
&= E(s \to t) \tag{28}
\end{aligned}
$$

where the final equality follows upon reversing the order of the first five equations with the roles of s and t exchanged.

Showing Eq. (26),

$$
[E(\tau_n \mid O = s)]^2 > r_{s,t}^2 [E(\tau_n \mid O = t) - E(\tau_n \mid O = s)]^2
$$

is equivalent to showing

$$
r_{s,t}[E(\tau_n \mid O = t) - E(\tau_n \mid O = s)] < [E(\tau_n \mid O = s)] \tag{29}
$$

As noted, this equation is valid only when the proposal probability $q_{s,t} > 0$ (so that the event O makes sense) and $r_{s,t}(\vec{\eta}) < 1$ (so that $E(\tau_n \mid O = s)$

makes sense). The steps in the derivation of Eq. (29), with an explanation to follow:

$$r_{s,t}[E(\tau_n \mid O=t) - E(\tau_n \mid O=s)]$$

$$= r_{s,t}E(t \to s)$$

$$= r_{s,t}E(s \to t)$$

$$< r_{s,t}E(s \to t, \text{last step is from } s)$$

$$= r_{s,t} \sum_{i=1}^{\infty} r_{s,t}(\vec{\eta})[1 - r_{s,t}(\vec{\eta})]^{i-1} iE(\tau_n \mid O=s)$$

$$= \frac{r_{s,t}(\vec{\eta})^2}{[1 - r_{s,t}(\vec{\eta})]} E(\tau_n \mid O=s) \sum_{i=1}^{\infty} i[1 - r_{s,t}(\vec{\eta})]^i$$

$$= E(\tau_n \mid O=s)$$

The first equality follows because the number of Metropolis steps required to reach an (s, t) proposed move starting from state t can be decomposed into

   • the number of Metropolis steps required to first reach state s when starting from state t, and

   • the number of Metropolis steps required to first propose an (s, t) move when starting in state s.

The latter event has distribution identical to $\{\tau_n \mid O=s\}$, and thus the term $E(\tau_n \mid O=t) - E(\tau_n \mid O=s)$ in Eq. (29) is simply equal to the expected number of steps for the simulated chain to move from state t to state s, written $E(t \to s)$. The second equation is simply Eq. (28). The third equation follows because the chain moves from s to t at least as fast as it moves from s to t with the last step being directly from s; this is because there exists a path from s to t that doesn't involve the final step being from s (otherwise, the proposal probabilities satisfy $q_{u,t}=0$ for all $u \neq s$, implying by symmetry that $q_{t,u}=0$ for all $u \neq s$, implying that $q_{t,s}=q_{s,t}=1$, implying that the chain is not irreducible and aperiodic). The fourth equation follows upon conditioning on the number i of (s, t) moves that are proposed before the one-step transition from s to t occurs, and the rest is algebra.

ACKNOWLEDGMENT

REFERENCES

1. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **22**:1087–1092 (1953).
2. M. Fitzgerald, R. R. Picard, and R. N. Silver, Canonical transition probabilities for adaptive Metropolis simulation, *Europhys. Lett.* **46**:282–287 (1999).
3. S. Chib and E. Greenberg, Understanding the Metropolis–Hastings algorithm, *American Statistician* **49**:327–335 (1995).
4. A. M. Ferrenberg, D. P. Landau, and R. H. Swendsen, Statistical errors in histogram reweighting, *Phys. Rev. E* **51**:5092–5100 (1995).
5. G. R. Smith and A. D. Bruce, A study of the multicanonical Monte Carlo method, *J. Phys. A: Math. Gen.* **28**:6623–6643 (1995).
6. E. Anderson *et al.*, SIAM (1994). [http://www.netlib.org/lapack/lug/lapack_lug.html].
7. J. S. Wang, T. K. Tay, and R. H. Swendsen, Transition matrix Monte Carlo reweighting and dynamics, *Phys. Rev. Lett.* **82**:476–479 (1999).
8. P. M. C. de Oliveira, T. J. P. Penna, and H. J. Herrmann, Broad histogram Monte Carlo, *Eur. Phys. J. B* **1**:205–208 (1998).
9. P. D. Beale, Exact distribution of energies in the two-dimensional Ising model, *Phys. Rev. Lett.* **76**:78–81 (1996).
10. B. A. Berg and T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions, *Phys. Rev. Lett.* **68**:9–12 (1992).
11. W. Feller, *An Introduction to Probability Theory and Its Application*, Vols. 1 and 2 (Wiley, New York, 1971).
12. P. H. Peskun, Optimum Monte Carlo sampling using Markov chains, *Biometrika* **60**:607–612 (1973).
13. V. I. Romanovsky, *Discrete Markov Chains* (Wolters-Noordhoff, Groningen, Netherlands, 1970).
14. C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, New York, 1973).