Towards a Definition of Equilibration for Markov Chains*

Robert D. Skeel
Department of Computer Science, Purdue University
West Lafayette, Indiana 47907-2107, U.S.A.

May 9, 2010

Abstract

Markov chain Monte Carlo methods are very popular for computing expectations. Their efficiency and reliability are subject to two significant drawbacks. The first is the correlation between successive samples. This reduces efficiency and frustrates variance estimation. The second drawback is the dependence on starting values, which leads to discarding a large initial set of "atypical" samples. The process of running the Monte Carlo method until getting an adequate starting value is called equilibration. Associated with this are two practical problems. One is to detect the onset of equilibration so that production may begin. The other is to characterize what it means to be equilibrated so that there might be a better understanding of how to initialize the equilibration process to reduce its running time. This article examines the statistical error of Monte Carlo method and proposes a definition of what it means to be equilibrated, which corresponds exactly to what is needed in practice and which is amenable to mathematical analysis.

1 Overview

Vast amounts of computer time are used to compute expectations of "observables" $A(\mathbf{x})$ that depend on random states $\mathbf{x}$ drawn from distributions that are known except for their normalizing factor. In particular, computing such averages constitutes the major part of calculations for applications described by statistical mechanics. Methods for doing so typically employ Markov chains constructed so that their stationary distribution is the target distribution. However, from a statistical point of view, there are two problems with this approach: The first is the high correlation between successive values generated by the Markov chain. This not only results in slow convergence but makes error bars much more difficult to compute. The second is the significant contamination introduced by poorly chosen starting values. In practice, this is handled by discarding samples until "equilibration" has been attained. The problem with this is not only the increased run time but the difficult question of deciding when to stop equilibration and start production.

To explore this latter problem satisfactorily requires a precise definition for what it means to be equilibrated. First, a distinction must be made between "equilibrated" and "in equilibrium." Being in equilibrium is necessarily a property of an ensemble $\mathbf{x}$ of values and implies that it has a stationary distribution $\rho(x)$. Being equilibrated appears to have no formal definition, but in common parlance the term is applied to a state x rather than an ensemble of states

*This material is based upon work supported by grant R01GM083605 from the National Institute of General Medical Sciences, and award A5286056128 from the University of Minnesota. Also, thanks to the Center for Biological Physics at Arizona State University and the Institute for Mathematics and Its Applications at the University of Minnesota for providing environments that facilitated this work.

x. In Section 2, after a review of Markov chain Monte Carlo (MCMC) methods, we propose a precise definition for being equilibrated that corresponds precisely to what is relevant in practice. The definition of being equilibrated is based on the following thought experiment: Imagine generating a very large number N of samples from state x and using all but the first m values to estimate the expectation. The state x is said to be equilibrated if when starting the Markov chain from that state, the optimal value of m is zero.

To explore these issues further, we obtain, in Section 3, an expression for the statistical error and its dependence on correlations and the starting value. The analysis is done for the case of a finite number of states. There is an excellent precedent for developing the details of a theory for Markov processes only for the case of a finite number of states and that is the landmark paper of Crooks [2]. Also reference [1, Chap. 4] assumes a finite number of states, which is justified by the use of finite-precision numbers (and the fact that finite precision seems to have no practical effect on the results). Extension to an infinite state space is possible, at least for a number of methods of interest.

For the most general case when the process is merely ergodic but not mixing, the effect of the starting value is very difficult to understand. So we make the additional assumption in Section 4 that the procedure is mixing, which is the normal situation for stochastic MCMCs. This permits mathematical analysis, which gives evidence that the definition of equilibrated is sensible.

The article closes with a discussion of the merit of the proposed definition. A good definition not only captures the desired property but is also reasonably simple and easy to apply. This, is, however, an open question. Another open question concerns a lower bound on the probability of a state being equilibrated if it is randomly chosen from the target distribution.

2 Markov Chain Monte Carlo Methods

Given some observable $A(x)$ defined for a set of states x, the problem is to estimate the expectation

$$\mu = \langle A(\mathbf{x}) \rangle = \int A(x)\rho(x)\mathrm{d}x$$

with respect to some probability density $\rho(x)$, which is known—except for its normalizing constant. A typical example is a Boltzmann-Gibbs distribution, for which the probability density is proportional to $\exp(-\beta E(x))$, where $E(x)$ is energy and β is the inverse temperature. The expectation is estimated by an average

$$\mu \approx \bar{\mathbf{A}}_{0,N} = \frac{1}{N}\sum_{n=0}^{N-1} A(\mathbf{x}_n)$$

where the $\mathbf{x}_n$ are random configurations, each with probability density $\rho(x)$. A typical value of N is one million.

Because the normalizing constant for $\rho(x)$ is generally very costly to compute, random configurations are generated using a method that does not require it, namely, a Markov chain Monte Carlo method. Such a method generates a sequence

$$\mathbf{x}_0 \to \mathbf{x}_1 \to \mathbf{x}_2 \to \cdots \to \mathbf{x}_{N-1},$$

where only in the limit $N \to \infty$ does the density $\rho_N(x)$ of $\mathbf{x}_N$ converge to $\rho(x)$—if the process is mixing. If the process is ergodic, it is the average $(1/N)\sum_{n=0}^{N-1} \rho_n(x)$ that converges to $\rho(x)$ in the limit $N \to \infty$.

The estimate *and its accuracy* depend on the initial value $\mathbf{x}_0$, and a poor choice will unduly contaminate the result. To avoid this, we first run the MCMC method starting with an arbitrary value $\mathbf{x}'_0$ until it yields a value $\mathbf{x}'_M$ which will not skew the estimate, a process called *equilibration*, and we then start production using $\mathbf{x}_0 = \mathbf{x}'_M$. To reduce the cost of equilibration M, a reasonably typical value is chosen for $\mathbf{x}'_0$.

As stated in the introduction, it is proposed that the suitability of a starting value x be evaluated by comparing with the result of dropping the first m values. Hence, introduce the more general average

$$\bar{\mathbf{A}}_{m,N}(x) \stackrel{\text{def}}{=} \frac{1}{N-m} \sum_{n=m}^{N-1} A(\mathbf{x}_n) \quad \text{given } \mathbf{x}_0 = x.$$

Using this notation, the dependence of the estimate $\bar{\mathbf{A}}_{0,N}$ on the initial state $\mathbf{x}_0$ can be made explicit via $\bar{\mathbf{A}}_{0,N} = \bar{\mathbf{A}}_{0,N}(\mathbf{x}_0)$.

As a measure of merit, we use

$$\mathbf{D}^2_{m,N}(x) \stackrel{\text{def}}{=} (\bar{\mathbf{A}}_{m,N}(x) - \mu)^2.$$

Its expectation includes both the bias and the variance of the estimate:

$$\left\langle \mathbf{D}^2_{m,N}(x) \right\rangle = (\langle \bar{\mathbf{A}}_{m,N}(x) \rangle - \mu)^2 + \mathrm{Var}[\bar{\mathbf{A}}_{m,N}(x)].$$

The number of samples needed to achieve a given error bar is (asymptotically) proportional to the variance of the estimate $\bar{\mathbf{A}}_{0,N}$. If the samples were uncorrelated, this variance would be σ^2/N where σ^2 is the variance of $A(\mathbf{x})$. This motivates the following measure of the effect of correlated samples, introduced in [4] and used elsewhere, e.g. [6].

Definition 1 *The* statistical inefficiency *of a Markov chain P w.r.t. an observable $A(x)$ is defined by*

$$\Phi = \lim_{N\to\infty} \frac{\mathrm{Var}[\bar{\mathbf{A}}_{0,N}(\mathbf{x}_0)]}{\sigma^2/N}$$

where σ^2 is the variance of $A(\mathbf{x})$.

Though the quantity Φ is typically greater than 1, it can be less than 1, as illustrated by Example 1 of Sec. 4.

A practical way of getting variance estimates is block averaging [3], though there is no way of getting statistically rigorous estimates without prior knowledge. In practice, prior knowledge might be provided in the form of a lower bound on N.

Though the effect of the starting state $\mathbf{x}_0$ is proportional to $1/N^2$ only, this nonetheless is considerable in practical applications. The effect on the $1/N^2$ term due to omitting the first m samples is the basis for the following definition:

Definition 2 *A state x is said to be* equilibrated *for a given Markov process with respect to an observable $A(x)$ if*

$$\limsup_{N\to\infty} N^2 \left\langle \mathbf{D}^2_{m,N}(x) - \mathbf{D}^2_{0,N}(x) \right\rangle \geq 0 \quad \textit{for } m \geq 1.$$

This definition of equilibrated differs from that in [5]. This new definition corresponds more closely to what matters in practice.

We defer further investigation of this property to Sec. 4 where an additional assumption is made.

A practical test for equilibration is the use of reverse cumulative averaging, in which cumulative averages are computed starting at the end of the chain. This process is terminated when the benefit of decreasing variance is outweighed by increasing bias due to including unequilibrated values near the beginning of the chain. This is illustrated in [6, Fig. 1(a)] (where in the caption, k' means g and Eq. (3) means Eq. (4)).

3 Ergodic Finite-State Markov Processes

Consider a finite set of states $x = 1, 2, \ldots, \nu$ and a Markov chain $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$ with transition probability matrix P, i.e.,

$$\Pr(\mathbf{x}_{n+1} = y \mid \mathbf{x}_n = x) = P_{xy}.$$

Then the probability distributions $\pi_{n,x} = \Pr(\mathbf{x}_n = x)$ satisfy the relation

$$\pi_{n+1}^{\mathsf{T}} = \pi_n^{\mathsf{T}} P.$$

It follows that P is a right stochastic matrix, meaning that $P \geq 0$ componentwise and $Pe = e$ where e is the vector of all ones. Since P completely characterizes the Markov chain, we identify P with the Markov chain.

Assume that P is irreducible. This means that there does *not* exist a permutation matrix J such that JPJ^{T} is block triangular with square diagonal blocks. From Perron-Frobenius theory, it follows that there is a unique left eigenvector satisfying $\pi^{\mathsf{T}} P = \pi^{\mathsf{T}}$ and $\pi^{\mathsf{T}} e = 1$, and, moreover, that $\pi^{\mathsf{T}} > 0$. The elements of π^{T} are the state probabilities for the unique stationary distribution of the Markov chain. We can write

$$P = e\pi^{\mathsf{T}} + Q,$$

where $Qe = 0$, $\pi^{\mathsf{T}} Q = 0^{\mathsf{T}}$, the spectral radius of Q is ≤ 1, and $\det(I - Q) \neq 0$. It follows that

$$P^k = e\pi^{\mathsf{T}} + Q^k, \quad k \geq 1.$$

Let A be an observable with values $A(x) = A_x$. Then $\langle A(\mathbf{x}_n)\rangle = \pi_n^{\mathsf{T}} \mathrm{diag}(A_1, A_2, \ldots, A_\nu) e$ (recalling that ν is the number of states), and the expectation for a stationary distribution

$$\mu = \langle A(\mathbf{x})\rangle = \pi^{\mathsf{T}} \mathrm{diag}(A_1, A_2, \ldots, A_\nu) e,$$

where $\mathbf{x}$ denotes a random state with state probabilities π^{T}. Define the deviation

$$D = \mathrm{diag}(A_1, A_2, \ldots, A_\nu) - \mu I,$$

and we have

$$\langle (A(\mathbf{x}_n) - \mu)(A(\mathbf{x}_{n+k}) - \mu)\rangle = \sum_x \sum_y (A_x - \mu)(A_y - \mu)\pi_{n,x}(P^k)_{xy} = \pi_n^{\mathsf{T}} D P^k D e. \qquad (1)$$

(Note that $\pi_{n,x}(P^k)_{xy}$ is a joint probability.) The variance of $A(\mathbf{x})$ is $\sigma^2 = \pi^{\mathsf{T}} D^2 e$.

The bias of an estimate due to a starting state x of the production run is given by the following:

Proposition 1

$$\langle \bar{\mathbf{A}}_{m,N}(x)\rangle = \mu + \frac{1}{N-m} e_x^{\mathsf{T}} Q^m (I - Q)^{-1} (I - Q^{N-m}) D e.$$

where e_x is the vector that has component x equal to 1 and the others set to zero.

Proof.

$$\langle \bar{\mathbf{A}}_{m,N}(x)\rangle = \frac{1}{N-m}\sum_{n=m}^{N-1}\langle A(\mathbf{x}_n)\rangle = \frac{1}{N-m}\sum_{n=m}^{N-1}\pi_n^{\mathsf{T}}(\mu I + D)e = \mu + \frac{1}{N-m}\sum_{n=m}^{N-1}\pi_n^{\mathsf{T}}De.$$

We have

$$\pi_n^{\mathsf{T}} = \pi_0^{\mathsf{T}}P^n = \pi^{\mathsf{T}} + \pi_0^{\mathsf{T}}Q^n, \quad \text{for } n \geq 1. \tag{2}$$

Hence,

$$\langle \bar{\mathbf{A}}_{m,N}(x)\rangle = \mu + \frac{1}{N-m}\sum_{n=m}^{N-1}\pi_0^{\mathsf{T}}Q^nDe.$$

□

We deduce from this that the value of m that minimizes the bias might be quite large. Of most interest is a combined effect of bias and variance, which is measured by $\langle \mathbf{D}^2_{m,N}(x)\rangle$.

For simplicity we state the next result for the case $m = 0$; the generalization is straightforward.

Proposition 2

$$\begin{aligned}\left\langle \mathbf{D}^2_{0,N}(x)\right\rangle &= \frac{\alpha}{N} + \frac{\beta_0(x)}{N^2} + \frac{2}{N^2}\pi^{\mathsf{T}}DQ^N(I-Q)^{-2}De - \frac{1}{N^2}e_x^{\mathsf{T}}Q^Nc \\ &\quad - \frac{2}{N^2}\sum_{n=0}^{N-1}e_x^{\mathsf{T}}Q^nDQ^{N-n}(I-Q)^{-1}De\end{aligned}$$

where

$$\begin{aligned}\alpha &= \pi^{\mathsf{T}}c, \\ \beta_0(x) &= \beta + e_x^{\mathsf{T}}c, \\ \beta &= \pi^{\mathsf{T}}D^2e - 2\pi^{\mathsf{T}}D(I-Q)^{-2}De, \\ c &= (I-Q)^{-1}D(I+Q)(I-Q)^{-1}De.\end{aligned}$$

Proof. Using Eq. (1),

$$\begin{aligned}\left\langle (\bar{\mathbf{A}}_{0,N}(x)-\mu)^2\right\rangle &= \frac{1}{N^2}\sum_{n=0}^{N-1}\left\langle (A(\mathbf{x}_n)-\mu)^2\right\rangle + \frac{2}{N^2}\sum_{n=0}^{N-2}\sum_{p=n+1}^{N-1}\langle (A(\mathbf{x}_n)-\mu)(A(\mathbf{x}_p)-\mu)\rangle \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\pi_n^{\mathsf{T}}D^2e + \frac{2}{N^2}\sum_{n=0}^{N-2}\pi_n^{\mathsf{T}}D(P+\cdots+P^{N-n-1})De \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\pi_n^{\mathsf{T}}D^2e + \frac{2}{N^2}\sum_{n=0}^{N-2}\pi_n^{\mathsf{T}}D(Q-Q^{N-n})(I-Q)^{-1}De,\end{aligned}$$

where the last equality uses Eq. (2) and the fact that $\pi^{\mathsf{T}}De = 0$. Using $\pi_n^{\mathsf{T}} = \pi^{\mathsf{T}} + \pi_0^{\mathsf{T}}Q^n$, $n \geq 1$, gives

$$\begin{aligned}\left\langle (\bar{\mathbf{A}}_{0,N}(x)-\mu)^2\right\rangle &= \frac{N-1}{N^2}\pi^{\mathsf{T}}D^2e + \frac{1}{N^2}\pi_0^{\mathsf{T}}(I-Q)^{-1}(I-Q^N)D^2e \\ &\quad + \frac{2(N-2)}{N^2}\pi^{\mathsf{T}}DQ(I-Q)^{-1}De - \frac{2}{N^2}\pi^{\mathsf{T}}D(Q^2-Q^N)(I-Q)^{-2}De \\ &\quad + \frac{2}{N^2}\pi_0^{\mathsf{T}}(I-Q^{N-1})(I-Q)^{-1}DQ(I-Q)^{-1}De \\ &\quad - \frac{2}{N^2}\sum_{n=0}^{N-2}\pi_0^{\mathsf{T}}Q^nDQ^{N-n}(I-Q)^{-1}De.\end{aligned}$$

Note the simplification $\pi^\mathsf{T} c = \pi^\mathsf{T} D(I+Q)(I-Q)^{-1}De$, and the proposition follows. $\square$

For the given error metric, the dependence on the Markov process is encapsulated in Q and the dependence on the observable in D.

4 Mixing Finite-State Markov Processes

We assume here that the Markov process is mixing, which means that π_n always converges to the stationary distribution π and implies that the spectral radius of Q is < 1.

Proposition 3

$$\left\langle \mathbf{D}^2_{m,N}(x) \right\rangle = \frac{\alpha}{N} + \frac{\beta_m(x)}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)$$

where

$$\beta_m(x) = m\alpha + \beta + e_x^\mathsf{T} P^m c.$$

Proof. From Proposition 2, we have

$$\left\langle \mathbf{D}^2_{0,N}(x) \right\rangle = \frac{\alpha}{N} + \frac{1}{N^2}\left(\beta + e_x^\mathsf{T} c\right) + \mathcal{O}\left(\frac{1}{N^3}\right).$$

Employing a conditional expectation, the "straightforward" generalization is

$$\left\langle \mathbf{D}^2_{m,N}(x) | \mathbf{x}_m = y \right\rangle = \frac{\alpha}{N-m} + \frac{1}{(N-m)^2}\left(\beta + e_y^\mathsf{T} c\right) + \mathcal{O}\left(\frac{1}{N^3}\right),$$

whence

$$\begin{aligned}
\left\langle \mathbf{D}^2_{m,N}(x) \right\rangle &= \sum_y \pi_{m,y} \left\langle \mathbf{D}^2_{m,N}(x) | \mathbf{x}_m = y \right\rangle \\
&= \frac{\alpha}{N-m} + \frac{1}{(N-m)^2}\left(\beta + \pi_m^\mathsf{T} c\right) + \mathcal{O}\left(\frac{1}{N^3}\right) \\
&= \frac{\alpha}{N} + \frac{1}{N^2}\left(m\alpha + \beta + e_x^\mathsf{T} P^m c\right) + \mathcal{O}\left(\frac{1}{N^3}\right).
\end{aligned}$$

$\square$

A short calculation based on Proposition 3 gives the statistical inefficiency:

$$\Phi = \frac{\alpha}{\sigma^2} = 1 + \frac{2\pi^\mathsf{T} DQ(I-Q)^{-1}De}{\pi^\mathsf{T} D^2 e}.$$

The following is an immediate consequence:

Corollary 1

$$\alpha = \pi^\mathsf{T} c \geq 0.$$

It may be helpful to have a more direct proof based solely on first principles:

Alternative proof. Defining $\Pi = \mathrm{diag}(\pi)$, we have

$$\pi^\mathsf{T} c = e^\mathsf{T} \Pi D(I+Q)(I-Q)^{-1}De = e^\mathsf{T} D(I-Q)^{-\mathsf{T}}(I-Q)^\mathsf{T}\Pi(I+Q)(I-Q)^{-1}De.$$

With $v = (I-Q)^{-1}De$, this becomes

$$\pi^\mathsf{T} c = v^\mathsf{T}(I-Q)^\mathsf{T}\Pi(I+Q)v = v^\mathsf{T}(I - Q^\mathsf{T}\Pi Q)v = v^\mathsf{T}(I - P^\mathsf{T}\Pi P)v = v^\mathsf{T}\Pi(I-\hat{P}P)v$$

where $\hat{P} = \Pi^{-1} P^{\mathsf{T}} \Pi$. The matrix $\hat{P}$ is a right stochastic matrix with stationary probability vector π^{T}. It gives the transition probability for what is called the *time reversal* of the process [2]. Hence the product $\hat{P}P$ is a right stochastic matrix for stationary probability vector π^{T}. With $w = \Pi^{1/2} v$, we have

$$\pi^{\mathsf{T}} c = w^{\mathsf{T}} (I - \Pi^{1/2} \hat{P} P \Pi^{-1/2}) w.$$

This must be nonnegative because $\Pi^{1/2} \hat{P} P \Pi^{-1/2}$ is a symmetric matrix similar to a matrix $\hat{P}P$ whose spectral radius is one. □

The proof of the following is short and straightforward:

Proposition 4 *State x is equilibrated if and only if*

$$\beta_0(x) \leq \beta_m(x) \quad \textit{for } m \geq 1.$$

This can be rewritten as

$$e_x^{\mathsf{T}} (I - P^m) c \leq m \pi^{\mathsf{T}} c \quad \textit{for } m \geq 1$$

or as

$$e_x^{\mathsf{T}} (I - Q^m) c \leq (m+1) \pi^{\mathsf{T}} c \quad \textit{for } m \geq 1. \tag{3}$$

Proof. We have from Proposition 3 that

$$\limsup_{N \to \infty} N^2 \left\langle \mathbf{D}_{m,N}^2(x) - \mathbf{D}_{0,N}^2(x) \right\rangle = \beta_m(x) - \beta_0(x).$$

□

The definition proposed in [5] requires that an equilibrated state x be "as good as random," meaning generally that $\beta_0(x)) \leq \langle \beta_0(\mathbf{x}) \rangle$. This is equivalent to $e_x^{\mathsf{T}} c \leq \pi^{\mathsf{T}} c$, which is perhaps more stringent than the requirement of Eq. (3).

The proof of the following is straightforward:

Proposition 5

$$\langle \beta_m(\mathbf{x}) - \beta_0(\mathbf{x}) \rangle = m\alpha \geq 0.$$

What this result means is that for each m there is a positive probability that condition (3) is satisfied by a random state $\mathbf{x}$, because the probability that a random variable exceeds or equals its expectation is positive.

Following is a concrete example:

Example 1 *Let the observable be given by $\mu I + D$ with*

$$D = \mathrm{diag}(1, -1, 0),$$

and the process by $P = e\pi^{\mathsf{T}} + Q$ with

$$\pi^{\mathsf{T}} = \begin{bmatrix} \frac{1-\varepsilon}{2} & \frac{1-\varepsilon}{2} & \varepsilon \end{bmatrix}, \quad Q = \frac{1}{2} \lambda D e e^{\mathsf{T}} D,$$

subject to the conditions $0 < \varepsilon < 1$ and $|\lambda| \leq 1 - \varepsilon$. The variance $\sigma^2 = \pi^{\mathsf{T}} D^2 e = 1 - \varepsilon$. Using the relation $QDe = \lambda De$, we get $c = (1+\lambda)(1-\lambda)^{-1} D^2 e$, $\alpha = \pi^{\mathsf{T}} c = (1-\varepsilon)(1+\lambda)/(1-\lambda)$, and the statistical inefficiency

$$\Phi = \frac{1+\lambda}{1-\lambda}.$$

Example 2 *(Continuation of example 1.) The definition of equilibrated, Eq. (3) simplifies to $e_x^{\mathsf{T}} c \leq (m+1)\pi^{\mathsf{T}} c$, for $m \geq 1$. Hence, we need only that*

$$e_x^{\mathsf{T}} D^2 e \leq 2(1-\varepsilon).$$

For $x = 1, 2$ this amounts to $\varepsilon \leq \frac{1}{2}$; whereas for $x = 3$, this always holds. Hence, state 3 is always equilibrated and states 1 and 2 only if their combined probability is at least one half. Therefore, the probability of a random state being equilibrated is always at least one half.

5 Discussion

The utility of the proposed definition of equilibrated is greatly enhanced if the following is true:

Open question 1 *Can the characterization of being equilibrated given by Proposition 4 be simplified? In particular, can it be shown that if inequality (3) holds for $m = 1$, it must also hold for $m \geq 2$? Or, at least, can a sufficient condition be obtained that does not involve enumerating the values $m = 1, 2, \ldots$?*

Additionally, it seems particularly difficult to simplify the equilibration condition unless the following is true:

Open question 2 *Is $\pi^{\mathsf{T}} c > 0$ if $c \neq 0$?*

It is unlikely that the approach taken to prove the corollary to Proposition 3 can be strengthened to affirm the second open question. Such an approach seems to require that the eigenvalue 1 of the matrix $\hat{P}P$ be of multiplicity 1 (where $\hat{P}$ is defined in the proof.) The example that follows shows that this need not be the case:

Example 3 *There exists an irreducible right stochastic matrix P for which the associated Markov process is mixing and yet $\hat{P}P$ is reducible: Let P be the 4 by 4 matrix $\frac{1}{4}ee^{\mathsf{T}} + \frac{1}{4}uv^{\mathsf{T}}$ with $u = \begin{bmatrix} 1 & -1 & 1 & -1 \end{bmatrix}^{\mathsf{T}}$ and $v = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^{\mathsf{T}}$. Then $\pi^{\mathsf{T}} = \frac{1}{4}e^{\mathsf{T}}$ and $\hat{P}P$ is block diagonal.*

Finally, intuition suggests that equilibrated states should be bountiful in the sense that if a state is chosen at random from the distribution π, the probability that it is equilibrated is not "too low":

Open question 3 *Is there a a lower bound on the probability that a random state is equilibrated, i.e., a positive number θ independent of m such that $\Pr(\beta_0(\mathbf{x}) \leq \beta_m(\mathbf{x})) \geq \theta$ for all $m \geq 1$? Is there such a bound that is independent of D, Q, and/or π, e.g., $\theta = \frac{1}{2}$?*

References

[1] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Clarendon Press, Oxford, New York, 1987. Reprinted in paperback in 1989 with corrections.

[2] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61(3):2361–2366, Mar 2000.

[3] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, second edition, 2002.

[4] R. Friedberg and J. E. Cameron. Test of the Monte Carlo method: Fast simulation of a small Ising lattice. *J. Chem. Phys.*, 52:6049–6058, 1970.

[5] R. D. Skeel. Two-point boundary value problems for curves: The case of minimum free energy paths. In T. E. Simos, G. Psihoyios, and C. Tsitouras, editors, *Numerical Analysis and Applied Mathematics: International Conference on Numerical Analysis and Applied Mathematics 2009*, volume 1168/1 of *AIP Conference Proceedings*, pages 29–31. Springer, 2009.

[6] W. Yang, R. Bitetti-Putzer, and M. Karplus. Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.*, 120(6):2618–2628, Feb. 8, 2004.