Elect. Comm. in Probab. **6** (2001) 79–89

ELECTRONIC COMMUNICATIONS in PROBABILITY

HOW TO COMBINE FAST HEURISTIC MARKOV CHAIN MONTE CARLO WITH SLOW EXACT SAMPLING

ANTAR BANDYOPADHYAY
University of California, Department of Statistics
367 Evans Hall # 3860, Berkeley CA 94720-3860, USA
email: antar@stat.berkeley.edu

DAVID J. ALDOUS[1]
University of California, Department of Statistics
367 Evans Hall # 3860, Berkeley CA 94720-3860, USA
email: aldous@stat.berkeley.edu

submitted June 6, 2001 *Final version accepted* July 28, 2001

AMS 2000 Subject classification: 60J10, 62M05, 68W20
Confidence interval, Exact sampling, Markov chain Monte Carlo

Abstract

Given a probability law π on a set S and a function $g : S \to R$, suppose one wants to estimate the mean $\bar{g} = \int g \, d\pi$. The Markov Chain Monte Carlo method consists of inventing and simulating a Markov chain with stationary distribution π. Typically one has no a priori *bounds on the chain's mixing time, so even if simulations suggest rapid mixing one cannot infer rigorous confidence intervals for $\bar{g}$. But suppose there is also a separate method which (slowly) gives samples exactly from π. Using n exact samples, one could immediately get a confidence interval of length $O(n^{-1/2})$. But one can do better. Use each exact sample as the initial state of a Markov chain, and run each of these n chains for m steps. We show how to construct confidence intervals which are always valid, and which, if the (unknown) relaxation time of the chain is sufficiently small relative to m/n, have length $O(n^{-1} \log n)$ with high probability.*

1 Background

Let π be a given probability distribution on a set S. Given a function $g : S \to R$, we want to estimate its mean $\bar{g} := \int_S g(s)\pi(ds)$. As we learn in elementary statistics, one can obtain an estimate for $\bar{g}$ by taking samples from π and using the sample average g-value as an estimator. But algorithms which sample exactly from π may be prohibitively slow. This is the setting for the *Markov chain Monte Carlo* (MCMC) method, classical in statistical physics and over the last ten years studied extensively as statistical methodology [4, 7, 9, 12]. In MCMC one designs a Markov chain on state-space S to have stationary distribution π. Then the sample average g-value over a long run of the chain is a heuristic estimator of $\bar{g}$. Diagnostics for assessing

[1]This material is based upon work supported by the National Science Foundation under Grant No. 9970901

length of run required, and expressions for heuristic confidence intervals, form a substantial part of MCMC methodology [11]. In general one cannot make such estimates rigorous, because one cannot eliminate the possibility that all the samples seen come from some small part of the state space which is almost disconnected from the remainder. Rigorous estimates typically require an a priori bound on some notion of the chain's *mixing time* (e.g. the relaxation time defined at (3)); and while there is now substantial theoretical literature on mixing times [1, 3, 5] it deals with settings more tractable than most statistical applications.
This paper investigates the interface between rigor and heuristics in a particular (perhaps artificial) context. Suppose we have a guess $\hat{\tau}$ for the mixing time of the chain, based on simulation diagnostics or heuristic estimates [6] or some non-rigorous mathematical argument. Suppose we have some separate scheme (an *exact sampler*) which gives independent samples exactly from π. Imagining that sampling $\hat{\tau}$ steps of the chain is roughly equivalent to sampling once from π, it is natural to consider the ratio

$$\rho = \frac{\text{cost of one exact sample}}{\text{cost of } \hat{\tau} \text{ steps of the chain}}$$

where *cost* refers to computational time. If $\rho < 1$ then one would just use the exact sampler and forget MCMC. If ρ is extremely large then we might not be able to afford even one exact sample, and we are forced to rely on MCMC (this is the setting typically envisaged in MCMC). This paper addresses the remaining context, where ρ is large but not extremely large; in other words, we can afford to simulate many steps of the chain (enough to make estimates heuristically good) but can afford only a few exact samples. In the case of sampling from general d-dimensional densities, for instance, exact samplers (e.g. based on rejection sampling using some tractable comparison density) are typically feasible only for small d, and MCMC is used for large d, so there should always be a window of d-values which fits our "ρ large, but not extremely large" context.
In this context, we could just use the exact sampler to get n independent samples from π. Then the sample average g-value provides an estimate of $\bar{g}$ with $O(n^{-1/2})$ error. But instead, suppose we use these n independent samples as initial states and generate n independent m-step realizations of the Markov chain. If diagnostic tests suggest that mixing occurs in $\hat{\tau}$ steps then we have an "effective sample size" of $(n \times m/\hat{\tau})$ and the heuristic estimate of error (when we use the overall sample average g-value as an estimator) would be $O(\sqrt{\hat{\tau}/(nm)})$. Our basic result, Theorem 2.1, shows that in a certain sense such error bounds can be made rigorous. This basic result requires an initial guess for mixing time; an "adaptive" variation, Theorem 2.2, eliminates that requirement.

2 Results

The discussion in section 1 provides conceptual context for our result, but let us now state the (rather minimal) mathematical assumptions for the result. For simplicity we assume the state space S is finite (though since our results are non-asymptotic they must extend to the general case without essential change). We assume (for reasons explained in Section 2.1) the function $g : S \to R$ is bounded, so by rescaling we may assume

$$0 \le g(\cdot) \le 1. \tag{1}$$

We assume the Markov chain is reversible, that is to say its transition matrix K satisfies

$$\pi_i k_{ij} = \pi_j k_{ji}, \quad \forall i, j. \tag{2}$$

With no further assumptions, Theorem 2.1 says we can construct a conservative confidence interval I which is always *valid*, i.e. satisfies (4). That is, for validity there are no "implicit asymptotics", and indeed we do not even need to assume K is irreducible. The *length* of the confidence interval will depend on the data, i.e. the realizations of the chain, but (5) bounds the length in terms of the *relaxation time* of the chain, defined as

$$\tau_2 := (1 - \lambda_2)^{-1} \tag{3}$$

where $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s \geq -1$ are the eigenvalues of K, and $\tau_2 < \infty$ if K is irreducible.

Theorem 2.1 *Assume (1,2). Take $n \geq 2$, $m \geq 1$, $0 < \alpha < 1$, $c > 0$ and $\hat{\tau} \geq 1$. Based on $2n$ exact samples from π and $2mn$ steps of the K-chain, we can construct an interval I such that*

$$\mathbf{P}\left(\bar{g} \notin I\right) \leq \alpha \tag{4}$$

and

$$\mathbf{P}\left(\mathit{length}(I) > k_\alpha^c \max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right) \log_2 n \right) \leq 3n(n+1)\exp\left(-\frac{m\, c^2}{48 n \tau_2} \log_2^2 n \right), \tag{5}$$

where $k_\alpha^c := 2\left(c\sqrt{2/\alpha} + \log(4/\alpha)\right)$.

2.1 Discussion

Of course $1 - \alpha$ is the prescribed confidence level. Choice of c reflects a trade-off between the "target length" $k_\alpha^c \max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right) \log_2 n$ of confidence interval and the probability of meeting that target. We have not tried to optimize numerical constants; at various places where we have used Chebyshev's inequality, more complicated arguments would presumably give better bounds.

To interpret Theorem 2.1 in order-of-magnitude terms, suppose we take $m = n\hat{\tau}$, so that we use $2n^2\hat{\tau}$ steps of the chain. Then the "target length" of our confidence interval will be $O(n^{-1}\log n)$. The theorem guarantees a confidence interval that is always valid, and guarantees that, if τ_2 is indeed not more than $\hat{\tau}$, then the length of the confidence interval will likely not exceed the target length. This contrasts with the $O(n^{-1/2})$ length confidence interval obtained by using the exact sampler only, and gets close to the $O(n^{-1})$ length of the heuristic confidence interval.

Admittedly the numerical constants make our numerical error bounds too crude to be useful in practice. For instance, taking $m = n\hat{\tau}$, $\alpha = 0.32$ and $c = 1$, the target length is about $10\frac{\log_2 n}{n}$, compared to the $\frac{1.0}{\sqrt{n}}$ length of the usual confidence interval based on exact samples only. So not until $n \approx 30,000$ would our interval be guaranteed to be actually shorter (provided $\hat{\tau}/\tau_2$ were sufficiently large). But the order-of-magnitude result does seem of theoretical interest, in particular because of its similarity to the idea [10] of self-testing algorithms. That paper describes an algorithm for generating random self-avoiding walks. As the authors write [10] "While there are a number of Monte Carlo algorithms used to solve these problems in practice, these are heuristic and their correctness relies on unproven conjectures. In contrast, our algorithm is shown rigorously to produce answers with specific accuracy and confidence. Only the efficiency of the algorithm relies on a widely believed conjecture, and a novel feature is that this conjecture can be *tested* as the algorithm proceeds." In our MCMC setting, we cannot estimate rigorously the actual value of τ_2, but we can self-justify inferences based on estimated τ_2.

On a more technical note, let us outline why Theorem 2.1 gives close to the best possible bounds on confidence interval length. Indeed, we claim that the best one could hope for is length of order

$$\max\left(\frac{1}{n}, \sqrt{\frac{\tau_2}{nm}}\right). \tag{6}$$

The point is that there are two different "obstacles" to sharp estimation. First, consider the eigenvector g_2 associated with eigenvalue λ_2. It is easy to estimate the variance of the overall sample average when $g = g_2$, and this variance works out as order $\frac{\tau_2}{nm}$; so we should not hope to have smaller estimation error than the corresponding standard deviation $\sqrt{\frac{\tau_2}{mn}}$. Second, suppose some subset A of the state space, with $\pi(A) = 1/n$, is almost disconnected. Then it is not unlikely that all n exact samples, and hence the n realizations of the chains, miss A, and so a contribution $E_\pi[g(\cdot)1_A]$ to $\bar{g}$ would be "invisible" to our simulations, and so this contribution is an unavoidable source of possible error when using sample averages as estimators. Our assumption (1) that g is bounded was intended as the simplest way of bounding this error – bounding it as order $1/n$.

So Theorem 2.1 shows that, if our initial guess $\hat{\tau}$ is indeed roughly close to τ_2, then our rigorous confidence interval's length will be roughly of the minimal order (6), up to $\log n$ terms. In Section 2.3 we give a natural "adaptive" variation in which we prescribe two numbers $\hat{\tau} < \hat{\tau}_{\max}$, where as before $\hat{\tau}$ is a heuristic estimate of τ_2, and where $2n^2\hat{\tau}_{\max}$ is the maximum number of steps of the chain that we would be willing to simulate. Theorem 2.2 gives an always-valid confidence interval which, if τ_2 is indeed small relative to $\hat{\tau}_{\max}$, will have length of order $n^{-1}\log n$ and will require order $n^2\tau_2$ steps of the chain.

2.2 Outline of construction and proof

Recall the "procedure" of simulating n realizations of m steps of the Markov chain, starting from n exact samples from π. The construction of the confidence interval I in Theorem 2.1 can be summarized as follows.

(i) Perform this procedure once, and find the overall average g-value – call it $\bar{g}^*$.

(ii) Perform the procedure again, and for $1 \le i \le n$ let A_i be the average g-value over the i'th m-step realization.

(iii) Test whether $|A_i - \bar{g}^*| \le \frac{c \log n}{\sqrt{r(n,m)}}$ for all i, where $r(n,m) := \min\left(n, \frac{m}{\hat{\tau}}\right)$. If so, report a "short" confidence interval $\left[\text{ave}_i A_i \pm O\left(\max\left(\frac{1}{n}, \sqrt{\frac{\hat{\tau}}{nm}}\right)\log n\right)\right]$; if not, report a "long" confidence interval $[\text{ave}_i A_i \pm O(\frac{1}{\sqrt{n}})]$.

To analyze the validity of the confidence interval, the key point is that after observing the event "$|A_i - \bar{g}^*| \le \frac{c \log n}{\sqrt{r(n,m)}}$" happening n times out of n, we can be confident that its probability is $1 - O(1/n)$. This allows us to truncate A_i at $\bar{g}^* \pm \frac{c \log n}{\sqrt{r(n,m)}}$, and then the sample average of n truncated variables has s.d. of order $\frac{\log n}{\sqrt{r(n,m)}} \times \frac{1}{\sqrt{n}} = \max\left(\frac{1}{n}, \sqrt{\frac{\hat{\tau}}{nm}}\right)\log n$.

Finally, to bound the chance of not reporting the short confidence interval we need to bound the chance of a truncation being needed, and a bound can be derived from large deviation estimates for reversible chains.

2.3 An adaptive version

In the procedure underlying Theorem 2.1 we make a single guess $\hat{\tau}$ and hope that the "good event" which leads to a short confidence interval will happen; if not, we settle for a long confidence interval. A natural variation is to specify that, if the "good event" fails, then repeat the process with $\hat{\tau}$ replaced by $2\hat{\tau}, 4\hat{\tau}, 8\hat{\tau}, \ldots$ and continue until the "good event" happens or until we reach some predetermined limit on numbers of steps of the chain.

Theorem 2.2 *Assume (1,2). Take $n \geq 5$, $0 < \alpha < 1$, and $1 \leq \hat{\tau} \leq \hat{\tau}_{\max} = 2^a \hat{\tau}$, where $a \geq 0$ is an integer. Then based on $2n$ exact samples from π, and $2n^2 \times M$ steps of the K-chain, where M is a random variable taking values in $\{\hat{\tau}, 2\hat{\tau}, 2^2\hat{\tau}, \ldots, 2^a\hat{\tau}\}$, we can define an interval I, such that*

$$\mathbf{P}\left(\bar{g} \notin I\right) \leq \alpha, \tag{7}$$

and

$$\textit{length}(I) \leq k_{\alpha,a} \frac{\log_2 n}{n} \quad \textit{if} \quad 1 \leq \frac{M}{\hat{\tau}} < 2^a, \tag{8}$$

and

$$\mathbf{P}\left(M > 96(\tau_2 \vee \hat{\tau})\right) \leq 3n(n+1)\exp\left(-\log_2^2 n\right). \tag{9}$$

where $k_{\alpha,a} = 2\left(\sqrt{2(a+1)/\alpha} + \log(4(a+1)/\alpha)\right)$.

So we are prescribing the maximum number of steps of the chain to be $2n^2\hat{\tau}_{\max}$. The bound in (9) is less than 0.05 for $n = 8$ and goes to zero very rapidly as n increases (the constant 96 just emerges from the proof technique). So if $\hat{\tau}_{\max}$ is indeed large compared to τ_2 then by generating a small number of exact samples one can construct a confidence interval for $\bar{g}$ which will be "short" with high probability, and the number of steps of the Markov chain required will be $O(n^2(\tau_2 \vee \hat{\tau}))$. More precisely, if we start with small $\hat{\tau}$ then we are confident of needing at most $192n^2\tau_2$ steps of the chain. But as discussed earlier, though the interval is "short" in the sense of being $O(\frac{\log n}{n})$, the numerical constants make the interval numerically large for moderate values of n.

3 Proof of Theorem 2.1

3.1 Construction of the confidence interval

Let $\{Z_1^*, Z_2^*, \ldots, Z_n^*\}$ be n independent samples from π. For $1 \leq i \leq n$ let $(X_{ij}^*)_{j=0}^{m-1}$ be a reversible Markov chain with initial state $X_{i0}^* = Z_i^*$; these n Markov chains are independent. Define

$$A_i^* := \frac{1}{m}\sum_{j=0}^{m-1} g(X_{ij}^*), \quad 1 \leq i \leq n,$$

and

$$\bar{g}^* := \frac{1}{n}\sum_{i=1}^{n} A_i^*. \tag{10}$$

$\bar{g}^*$ is our initial guess for $\bar{g}$.

Now re-run the entire simulation independently to get $\{Z_1, Z_2, \ldots, Z_n\}$, another set of n independent samples from π, and $(X_{ij})_{j=0}^{m-1}$ another independent but identically distributed family of n reversible Markov chains each with initial state $X_{i0} = Z_i$. We further define

$$A_i := \frac{1}{m}\sum_{j=0}^{m-1} g(X_{ij}), \quad 1 \le i \le n,$$

and

$$\overline{\overline{A}} := \frac{1}{n}\sum_{i=1}^{n} A_i. \tag{11}$$

Fix $c > 0$. Truncate each A_i to get

$$\tilde{A}_i := \begin{cases} A_i & \text{if} \quad |A_i - \bar{g}^*| \le c\,\frac{\log_2 n}{\sqrt{r(n,m)}} \\ \bar{g}^* & \text{otherwise} \end{cases} \tag{12}$$

where $r(n,m) := \min\left(n, \frac{m}{\hat{\tau}}\right)$. Let

$$\overline{\overline{\tilde{A}}} := \frac{1}{n}\sum_{i=1}^{n} \tilde{A}_i. \tag{13}$$

Write $N_n = \sum_{i=1}^{n} I(A_i \ne \tilde{A}_i)$ for the number of truncations, and call the event $G_n := [N_n = 0]$ the *good event.*
Define $h : (\mathbb{N} \cup \{0\}) \times \mathbb{N} \times (0,1) \to [0,\infty)$ by

$$h(z,n;\alpha) := \begin{cases} \dfrac{z}{n} + \dfrac{c_\alpha}{\sqrt{n}} & \text{if} \quad z \ne 0 \\ \dfrac{d_\alpha}{n} & \text{if} \quad z = 0 \end{cases}, \tag{14}$$

where

$$c_\alpha := \frac{1}{\sqrt{2\alpha}} \quad \text{and} \quad d_\alpha := \log\frac{2}{\alpha}.$$

In the next section we shall prove

Proposition 3.1 *For any $b > 0$*

$$\mathbf{P}\left[|\,\overline{\overline{\tilde{A}}} - \bar{g}| > b\,c\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right)\,\log_2 n + h(N_n, n; \alpha)\right] \le \frac{1}{b^2} + \alpha. \tag{15}$$

Replacing α by $\alpha/2$ in Proposition 3.1 and setting $b = \sqrt{2/\alpha}$, we see that the confidence interval

$$I := \overline{\overline{\tilde{A}}} \pm \left(c\sqrt{\frac{2}{\alpha}}\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right)\,\log_2 n + h(N_n, n; \alpha/2)\right) \tag{16}$$

satisfies the requirement of (4) that $\mathbf{P}(\bar{g} \notin I) \le \alpha$. If $N_n = 0$ then the length of this confidence interval is

$$2\left(c\sqrt{\frac{2}{\alpha}}\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right)\,\log_2 n + \frac{d_{\alpha/2}}{n}\right). \tag{17}$$

Notice that (17) is bounded by $k_\alpha^c \max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right) \log_2 n$; here we use assumption $n \ge 2$ which implies $\log_2 n \ge 1$. So to prove (5) and complete the proof of Theorem 2.1, it is enough to prove (in section 3.3)

Proposition 3.2

$$\mathbf{P}(N_n > 0) \le 3n(n+1)\exp\left(-\frac{m\,c^2}{48n\tau_2}\log_2^2 n\right).$$

3.2 Proof of Proposition 3.1

We denote the conditional probability, conditional expectation and conditional variance given $\bar{g}^*$ by $\mathbf{P}^{\bar{g}^*}$, $\mathbf{E}^{\bar{g}^*}$ and $\mathbf{Var}^{\bar{g}^*}$ respectively.

Observe that under $\mathbf{P}^{\bar{g}^*}$, the random variables $\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_n$ are i.i.d. Thus $\mathbf{E}^{\bar{g}^*}(\overline{\tilde{A}}) = \mathbf{E}^{\bar{g}^*}(\tilde{A}_1)$, and $\mathbf{Var}^{\bar{g}^*}(\overline{\tilde{A}}) = \frac{1}{n}\mathbf{Var}^{\bar{g}^*}(\tilde{A}_1)$. But $\mathbf{Var}^{\bar{g}^*}(\tilde{A}_1) = \mathbf{Var}^{\bar{g}^*}(\tilde{A}_1 - \bar{g}^*) \le \left(c\,\frac{\log_2 n}{\sqrt{r(n,m)}}\right)^2$, because $|\tilde{A}_1 - \bar{g}^*| \le c\,\frac{\log_2 n}{\sqrt{r(n,m)}}$. So by Chebyshev's inequality, we get

$$\mathbf{P}^{\bar{g}^*}\left[|\,\overline{\tilde{A}} - \mathbf{E}^{\bar{g}^*}(\tilde{A}_1)| > b\,c\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right)\log_2 n\right] \le \frac{1}{b^2},$$

and by taking expectation

$$\mathbf{P}\left[|\,\overline{\tilde{A}} - \mathbf{E}^{\bar{g}^*}(\tilde{A}_1)| > b\,c\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat{\tau}}{nm}}\right)\log_2 n\right] \le \frac{1}{b^2}. \tag{18}$$

Now we want to estimate $|\mathbf{E}^{\bar{g}^*}(\tilde{A}_1) - \bar{g}|$. ¿From the definitions (11) and (13),

$$(\overline{\overline{A}} - \overline{\tilde{A}}) = \frac{1}{n}\sum_{i=1}^{n} A_i I(A_i \ne \tilde{A}_i) - \frac{1}{n}\sum_{i=1}^{n} \bar{g}^* I(A_i \ne \tilde{A}_i).$$

Since g takes values in $[0,1]$,

$$-\frac{1}{n}N_n \;\le\; \left(\overline{\overline{A}} - \overline{\tilde{A}}\right) \;\le\; \frac{1}{n}N_n. \tag{19}$$

Now $\overline{\overline{A}}$ is independent of $\bar{g}^*$, thus taking conditional expectation given $\bar{g}^*$ in (19) we get

$$\left|\mathbf{E}^{\bar{g}^*}(\tilde{A}_1) - \bar{g}\right| \;\le\; p_n(\bar{g}^*), \tag{20}$$

where $p_n(\bar{g}^*) := \mathbf{P}^{\bar{g}^*}\left(|A_1 - \bar{g}^*| > c\log_2 n/\sqrt{r(n,m)}\right)$.

Under $\mathbf{P}^{\bar{g}^*}$ we have $N_n \sim \mathbf{Binomial}(n, p_n(\bar{g}^*))$, and hence

$$\begin{aligned}
\mathbf{P}^{\bar{g}^*}\left(p_n(\bar{g}^*) > \frac{d_\alpha}{n}, N_n = 0\right) &= \left(1 - p_n(\bar{g}^*)\right)^n 1_{(p_n(\bar{g}^*) > d_\alpha/n)} \\
&\le \left(1 - \frac{d_\alpha}{n}\right)^n 1_{(d_\alpha/n \le 1)} \\
&\le e^{-d_\alpha} \quad \text{since} \quad (1-x) \le e^{-x} \quad \forall\, 0 \le x \le 1 \\
&= \frac{\alpha}{2} \quad \text{by definition of} \quad d_\alpha.
\end{aligned} \tag{21}$$

Further,

$$
\begin{aligned}
&\mathbf{P}^{\bar g^*}\left(p_n(\bar g^*) > \frac{N_n}{n} + \frac{c_\alpha}{\sqrt{n}}, N_n > 0\right) && (22)\\
&\leq \mathbf{P}^{\bar g^*}\left(\frac{N_n}{n} + \frac{c_\alpha}{\sqrt{n}} < p_n(\bar g^*)\right)\\
&\leq \mathbf{P}^{\bar g^*}\left(|N_n - n p_n(\bar g^*)| > \sqrt{n} c_\alpha\right)\\
&\leq \frac{p_n(\bar g^*)(1-p_n(\bar g^*))}{c_\alpha^2} \text{ by Chebyshev's inequality}\\
&\leq \frac{1}{4c_\alpha^2}\\
&= \frac{\alpha}{2} \text{ by definition of } c_\alpha. && (23)
\end{aligned}
$$

Taking expectations of the conditional probabilities in (21) and (23) we get

$$\mathbf{P}\left(p_n(\bar g^*) > \frac{d_\alpha}{n}, N_n = 0\right) \leq \frac{\alpha}{2}$$

and

$$\mathbf{P}\left(p_n(\bar g^*) > \frac{N_n}{n} + \frac{c_\alpha}{\sqrt{n}}, N_n > 0\right) \leq \frac{\alpha}{2}.$$

Thus by definition of $h(\cdot)$

$$\mathbf{P}\Big(p_n(\bar g^*) > h(N_n, n; \alpha)\Big) \leq \alpha. \tag{24}$$

And hence from (18), (20) and (24) we get

$$
\begin{aligned}
&\mathbf{P}\left(|\,\tilde{\tilde{A}} - \bar g| > b\,c\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat\tau}{nm}}\right)\,\log_2 n + h(N_n, n; \alpha)\right)\\
\leq\ &\mathbf{P}\left(|\,\tilde{\tilde{A}} - \mathbf{E}(\tilde A_1|\bar g^*)| > b\,c\,\max\left(\tfrac{1}{n}, \sqrt{\tfrac{\hat\tau}{nm}}\right)\,\log_2 n\right)\\
&\qquad + \mathbf{P}\Big(p_n(\bar g^*) > h(N_n, n; \alpha)\Big)\\
\leq\ &\frac{1}{b^2} + \alpha.
\end{aligned}
$$

□

3.3 Proof of Proposition 3.2

Clearly

$$
\begin{aligned}
\mathbf{P}(N_n > 0) &\leq n\mathbf{P}\left(|A_1 - \bar g^*| > c\,\frac{\log_2 n}{\sqrt{r(n,m)}}\right)\\
&\leq n\left[\mathbf{P}\left(|A_1 - \bar g| > c\,\frac{\log_2 n}{2\sqrt{r(n,m)}}\right)\right.\\
&\qquad\left. + \mathbf{P}\left(|\bar g^* - \bar g| > c\,\frac{\log_2 n}{2\sqrt{r(n,m)}}\right)\right]. && (25)
\end{aligned}
$$

To bound the terms of (25) we use a large deviation bound for sample averages of reversible Markov chains. Lezaud [8] equation (2) gives a one-sided bound for $A_1 = m^{-1}\sum_{j=0}^{m-1} g(X_{1j})$:

$$\mathbf{P}\Big(A_1 - \bar{g} > \lambda\Big) \le \exp\left(\frac{1}{5\tau_2} - \frac{\lambda^2 m}{12\tau_2}\right), \quad \lambda > 0.$$

Since $\tau_2 \ge 1/2$ always, and $2e^{2/5} < 3$, we deduce the two-sided bound

$$\mathbf{P}\Big(|A_1 - \bar{g}| > \lambda\Big) \le 3\exp\left(-\frac{\lambda^2 m}{12\tau_2}\right), \quad \lambda > 0. \tag{26}$$

So in particular

$$\begin{aligned}\mathbf{P}\left(|A_1 - \bar{g}| > c\,\frac{\log_2 n}{2\sqrt{r(n,m)}}\right) &\le 3\exp\left(-\frac{m\,c^2}{48n\tau_2}\,\frac{n}{r(n,m)}\log_2^2 n\right)\\ &\le 3\exp\left(-\frac{m\,c^2}{48n\tau_2}\log_2^2 n\right).\end{aligned} \tag{27}$$

Also, for $\lambda > 0$,

$$\begin{aligned}\mathbf{P}\left(|\bar{g}^* - \bar{g}| > \lambda\right) &\le n\mathbf{P}\left(|A_1^* - \bar{g}| > \lambda\right)\\ &= n\mathbf{P}\left(|A_1 - \bar{g}| > \lambda\right)\\ &\le 3n\exp\left(-\frac{\lambda^2 m}{12\tau_2}\right) \text{ by (26).}\end{aligned}$$

So in particular

$$\begin{aligned}\mathbf{P}\left(|\bar{g}^* - \bar{g}| > c\,\frac{\log_2 n}{2\sqrt{r(n,m)}}\right) &\le 3n\exp\left(-\frac{m\,c^2}{48n\tau_2}\,\frac{n}{r(n,m)}\log_2^2 n\right)\\ &\le 3n\exp\left(-\frac{m\,c^2}{48n\tau_2}\log_2^2 n\right).\end{aligned} \tag{28}$$

Substituting (27) and (28) into (25) gives the bound asserted in Proposition 3.2.

□

4 Proof of Theorem 2.2

For the first part of the procedure for constructing the confidence interval I, simulate $\{Z_i^*, 1 \le i \le n\}$ and $\{Z_i, 1 \le i \le n\}$ as at the start of section 3.1. This part of the procedure is not repeated. Then for $k \in \{0, 1, 2, \ldots, a\}$, let $\{X_{ij}^* \,|\, 1 \le i \le n,\, 1 \le j \le m_k := 2^k n\hat{\tau}\}$ be realizations of chains started at $X_{i0}^* = Z_i^*$; and let $\{X_{ij} \,|\, 1 \le i \le n,\, 1 \le j \le m_k := 2^k n\hat{\tau}\}$ be realizations of chains started at $X_{i0} = Z_i$; each simulated until time $m_k := 2^k n\hat{\tau}$. Repeat definitions from Section 3: for $k = 0, 1, \ldots a$ define $A_i^{(k)}$, $\tilde{A}_i^{(k)}$, and $\overset{\simeq}{A}{}^{(k)}$ as A_i, $\tilde{A}_i$, and $\overset{\simeq}{A}$ respectively with $m = m_k$. Note that for such m we have $r(n,m) = n$.
Let $N_n^{(k)} = \sum_{i=1}^n I(A_i^{(k)} \ne \tilde{A}_i^{(k)})$. Define

$$I\,(n,k;\alpha) := \left[\overset{\simeq}{A}{}^{(k)} \pm \left(\sqrt{\frac{2}{\alpha}}\,\frac{\log_2 n}{n} + h(N_n^{(k)}, n; \alpha/2)\right)\right], \tag{29}$$

where $h(\cdot)$ is as defined in (14). This is the interval I defined at (16) which featured in Theorem 2.1 for $c = 1$, and so by (4) we get that for $0 \leq k \leq a$, $0 < \alpha < 1$, and $n \geq 5$,

$$\mathbf{P}\left(\bar{g} \notin I(n,k;\alpha)\right) \leq \alpha. \tag{30}$$

Define $T := \min\{\, 0 \leq k \leq a \,|\, N_n^{(k)} = 0 \,\}$, where we write $T = a$ if the set is empty. Then define

$$M := 2^T \hat{\tau} \in \{\hat{\tau}, 2\hat{\tau}, 2^2\hat{\tau}, \ldots, 2^a\hat{\tau}\}$$

.

$$I := I\Big(n, T; \alpha/(a+1)\Big).$$

If $0 \leq T < a$ then $N_n^{(T)} = 0$, and hence from (29) and (14) we get that

$$\mathrm{length}(I) \leq k_{\alpha,a}\, \frac{\log_2 n}{n}, \tag{31}$$

where $k_{\alpha,a} = 2\left(\sqrt{2(a+1)/\alpha} + \log(4(a+1)/\alpha)\right)$, so (8) is satisfied. Further,

$$\begin{aligned}
\mathbf{P}\Big(\bar{g} \notin I\Big) &= \sum_{k=0}^{a} \mathbf{P}\Big(\bar{g} \notin I, T = k\Big) \\
&= \sum_{k=0}^{a} \mathbf{P}\Big(\bar{g} \notin I(n,k;\alpha/(a+1)), T = k\Big) \\
&\leq \sum_{k=0}^{a} \mathbf{P}\Big(\bar{g} \notin I\left(n,k;\alpha/(a+1)\right)\Big) \\
&\leq \sum_{k=0}^{a} \frac{\alpha}{a+1} = \alpha.
\end{aligned}$$

So (7) is satisfied also.
Now to complete the proof we observe that

$$\begin{aligned}
\mathbf{P}\left(M > 96(\hat{\tau} \vee \tau_2)\right) &= \mathbf{P}\left(T > \left\lfloor \log_2 \frac{96(\tau_2 \vee \hat{\tau})}{\hat{\tau}} \right\rfloor\right) \\
&\leq \mathbf{P}\left(N_n^{(\lfloor \log_2 \frac{96(\tau_2 \vee \hat{\tau})}{\hat{\tau}} \rfloor)} > 0\right),
\end{aligned} \tag{32}$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.
Applying Proposition 3.2 with $m = n \times 2^{\lfloor \log_2 \frac{96(\tau_2 \vee \hat{\tau})}{\hat{\tau}} \rfloor} \hat{\tau}$, and $c = 1$, we get

$$\begin{aligned}
\mathbf{P}\left(N_n^{(\lfloor \log_2 \frac{96(\tau_2 \vee \hat{\tau})}{\hat{\tau}} \rfloor)} > 0\right) &\leq 3n(n+1)\exp\left(-\frac{2^{\lfloor \log_2 \frac{96(\hat{\tau} \vee \tau_2)}{\hat{\tau}} \rfloor}\, \hat{\tau}}{48\tau_2} \log_2^2 n\right) \\
&\leq 3n(n+1)\exp\left(-\log_2^2 n\right).
\end{aligned} \tag{33}$$

The bound asserted in (9) follows from (32) and (33). The number of chain steps used equals $2n^2 \times 2^T \hat{\tau} = 2n^2 \times M$.

$\square$

References

[1] D.J. Aldous and J.A. Fill. Reversible Markov chains and random walks on graphs. Book in preparation, 2001.

[2] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley, 1992.

[3] E. Behrends. *Introduction to Markov Chains, with special emphasis on rapid mixing.* Veiweg, 2000.

[4] M.-H. Chen, Q.-M. Shao, and J.G. Ibrahim. *Monte Carlo Methods in Bayesian Computation.* Springer–Verlag, 2000.

[5] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57:20–36, 1998.

[6] S.T. Garren and R.L. Smith. Estimating the second largest eigenvalue of a Markov transition matrix. *Bernoulli* 6: 215–242, 2000.

[7] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*, London, 1996. Chapman and Hall.

[8] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8:849–867, 1998.

[9] J.S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, 2001.

[10] D. Randall and A. Sinclair. Self-testing algorithms for self-avoiding walks. *J. Math. Phys.*, 41:1570–1584, 2000.

[11] C.P. Robert, editor. *Discretization and MCMC Convergence Assessment.* Number 135 in Lecture Notes in Statistics. Springer–Verlag, 1998.

[12] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer–Verlag, 1999.