MCMC Confidence Intervals and Biases

by (in alphabetical order)

Yu Hang Jiang, Tong Liu, Zhiya Lou, Jeffrey S. Rosenthal, Shanshan Shangguan, Fei Wang, and Zixuan Wu

Department of Statistical Sciences, University of Toronto

(December, 2020)

1 Introduction

Markov chain Monte Carlo (MCMC) is very a powerful tool for estimating and sampling from complicated high-dimensional distributions (see e.g. [2] and the many references therein). MCMC algorithms help researchers in a wide spectrum of fields, ranging from Bayesian statistics to finance to computer science to physics.

One of the biggest challenges when implementing MCMC algorithms is to evaluate the error of the estimate, which is crucial for generating accurate results, and can also help when deciding how many iterations of the chain should be run. The majority of the existing results for quantifying MCMC accuracy rely heavily on the Markov chain Central Limit Theorem (CLT). However, this CLT is only known to be valid under specific conditions like geometric ergodicity or reversibility, which do not always hold and can be difficult to verify (see e.g. [6, 7, 9, 11]). In the reversible case, Kipnis and Varadhan [10] established the existence of a CLT for all reversible Markov chains which have finite asymptotic estimator variance. However, without reversibility, CLTs are more challenging. Toth [15] generalized the results from [10] to the non-reversible case, but only under additional conditions which are very difficult to verify. And, Häggström [5] provides an example which shows that CLTs might not exist for non-reversible chains under conditions where CLTs would be guaranteed in the reversible case.

The recent paper [14] derived a simple MCMC confidence interval which does not require a CLT, using only Chebychev's inequality. That result required certain assumptions about how the estimator bias and variance grow with the number of iterations n, in particular that the bias is $o(1/\sqrt{n})$. This assumption seemed mild, since it is generally believed that the estimator bias will be $O(1/n)$ and hence $o(1/\sqrt{n})$ (see e.g. page 21 of [4]). However, questions were raised [1] about how to verify this assumption, and indeed we show herein (Section 4) that it might not always hold.

The present paper therefore seeks to simplify and weaken the assumptions in [14], to make MCMC confidence intervals without CLTs more widely applicable. In Section 2 (and Section 5), we derive an asymptotic Markov chain confidence interval assuming only a finite asymptotic estimator variance as in [10], without requiring any bias assumption nor reversibility nor stationarity nor a CLT; at significance level $\alpha = 0.05$ it is just 2.3 times as wide as the confidence interval that would follow from a CLT. In Section 3, we instead fix the number of iterations n, and obtain corresponding non-asymptotic confidence intervals without CLT under slightly stronger assumptions. In Section 4, we consider the question of when the MCMC bias is or is not $o(1/\sqrt{n})$, and show that this property does not always hold but is ensured by a polynomial ergodicity condition.

2 Asymptotic MCMC Confidence Intervals

Let $\{X_n\}$ be a ϕ-irreducible ergodic Markov chain on the state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$. Let $h : \mathcal{X} \to \mathbb{R}$ be a measurable function, let $\pi(h) = \int_{x \in \mathcal{X}} h(x)\, \pi(x)\, dx$ be the (finite) expression we wish to estimate, and let $e_n := \frac{1}{n} \sum_{i=1}^{n} h(X_i)$ be our estimate. At significance level α, we wish to find a conservative $1 - \alpha$ confidence interval, i.e. an interval which contains $\pi(h)$ with probability at least $1 - \alpha$. Using only a variance bound, we show the following:

Theorem 1: If $\limsup_{n\to\infty}\ n\, Var(e_n) \le B^2$ for some $B > 0$, then for any $0 < \alpha < 1$ and $\epsilon > 0$, and π-a.e. initial state $X_0 = x \in \mathcal{X}$, the interval

$$I_n := \big(e_n - (1+\epsilon) n^{-1/2} \alpha^{-1/2} B,\ e_n + (1+\epsilon) n^{-1/2} \alpha^{-1/2} B\big)$$

is an asymptotic conservative $1 - \alpha$ confidence interval for $\pi(h)$, i.e.

$$\liminf_{n\to\infty} P[\pi(h) \in I_n] \ \ge\ 1 - \alpha.$$

Proof. First assume the chain starts in stationarity, so $E(e_n) = \pi(h)$ for all $n \in \mathbb{N}$. Then for any $a_n > 0$, we have from Chebychev's inequality that

$$P\big(|e_n - \pi(h)| \ge a_n\big) \ = \ P\big(|e_n - E(e_n)| \ge a_n\big) \ \le \ Var(e_n) \ / \ a_n^2.$$

Therefore, setting $a_n = B/\sqrt{n\alpha} > 0$ gives

$$\begin{aligned}\limsup_{n\to\infty} P\big(|e_n - \pi(h)| \ge a_n\big) &\le \limsup_{n\to\infty} \big(Var(e_n)/a_n^2\big) \\ &\le \limsup_{n\to\infty} \big(Var(e_n) \frac{n\alpha}{B^2}\big) \\ &\le \limsup_{n\to\infty} (B^2 \frac{\alpha}{B^2}) = \alpha.\end{aligned}$$

Then, taking complements gives

$$\liminf_{n\to\infty} P\big(|e_n - \pi(h)| < n^{-1/2}\alpha^{-1/2}B\big) = \liminf_{n\to\infty} P\big(|e_n - \pi(h)| < a_n\big) \geq 1-\alpha.$$

This proves the result (with $\epsilon = 0$) assuming the chain starts in stationarity.

Finally, applying Theorem 5 from Section 5 below, with $\epsilon > 0$ and $r = 1/2$ and $C = \epsilon\,\alpha^{-1/2}B$, we obtain the result for π-a.e. $X_0 = x \in \mathcal{X}$. □

Theorem 1 says that any Markov chain satisfying $\limsup_{n\to\infty} n\,Var(e_n) \leq B^2$ for some $B > 0$ immediately has a specified asymptotic confidence interval, without requiring any CLT. It does not require any bias bound, so it provides a partial response to [1]. It does still require a variance bound. But asymptotic variance estimators can be obtained in many different ways, including repeated runs, integrated autocorrelation times, batch means, window estimators, regenerations, and more (see e.g. Section 3 of [3]), so this does not appear to be a major limitation.

For example, at the usual significance level $\alpha = 0.05$, taking $\epsilon = 0.001$, Theorem 1 yields the asymptotic 95% confidence interval

$$(e_n - 4.48\,B/\sqrt{n},\ e_n + 4.48\,B/\sqrt{n}).$$

By contrast, if we knew that a CLT held and that $\lim_{n\to\infty} nVar(e_n) = B^2$, then we could derive the 95% confidence interval

$$(e_n - 1.96\,B/\sqrt{n},\ e_n + 1.96\,B/\sqrt{n}).$$

The width of the first confidence interval is 2.3 times the second, but it does not require reversibility, nor the actual convergence of $n\,Var(e_n)$ as $n \to \infty$.

3 Non-asymptotic MCMC Confidence Intervals

The confidence intervals from Theorem 1 are only valid asymptotically as $n \to \infty$. That limitation is quite common for most MCMC confidence intervals, since large n is required for a CLT to hold. However, since we are not using any CLT in our analysis, it is possible to obtain a precise non-asymptotic interval, in terms of an upper bound on the bias, as follows.

Theorem 2: Suppose for some fixed $n \in \mathbb{N}$, the chain satisfies the variance bound $n\,Var(e_n) \leq B^2$ for some $B > 0$, and also the bias bound $|E(e_n) - \pi(h)| \leq C$ for some $C \geq 0$. Then for any significance level $\alpha \in (0,1)$, setting $\delta = \frac{C}{\frac{B}{\sqrt{n\alpha}}+C} \in [0,1)$ and $a_n = \frac{B}{\sqrt{n\alpha(1-\delta)}}$, the fixed-$n$ interval

$$I_n \ := \ (e_n - a_n,\ e_n + a_n)$$

is a non-asymptotic conservative $1-\alpha$ confidence interval, i.e.

$$P[\pi(h) \in I_n] \;\geq\; 1-\alpha, \qquad n \in \mathbb{N}.$$

Proof. We first compute that

$$\frac{\delta}{1-\delta} \;=\; \frac{C}{\frac{B}{\sqrt{n\alpha}}+C} \Bigg/ \frac{\frac{B}{\sqrt{n\alpha}}}{\frac{B}{\sqrt{n\alpha}}+C} \;=\; \frac{C\sqrt{n\alpha}}{B},$$

and hence

$$\delta a_n \;=\; \frac{\delta}{1-\delta}\,\frac{B}{\sqrt{n\alpha}} \;=\; \frac{C\sqrt{n\alpha}}{B}\,\frac{B}{\sqrt{n\alpha}} \;=\; C.$$

Thus $|E(e_n - \pi(h))| \leq C = \delta a_n$, and hence $a_n - |E(e_n) - \pi(h)| \geq a_n - C = (1-\delta)a_n > 0$. Therefore, using the triangle inequality and then Chebyshev's inequality, we have that

$$\begin{aligned} P\Big(|e_n - \pi(h)| \geq a_n|\Big) &= P\Big(|e_n - E(e_n) + E(e_n) - \pi(h)| \geq a_n\Big) \\ &\leq P\Big(|e_n - E(e_n)| + |E(e_n) - \pi(h)| \geq a_n\Big) \\ &= P\Big(|e_n - E(e_n)| \geq a_n - |E(e_n) - \pi(h)|\Big) \\ &\leq Var(e_n) \;/\; \big(a_n - |E(e_n) - \pi(h)|\big)^2 \\ &\leq Var(e_n) \;/\; \big((1-\delta)a_n\big)^2 \\ &= Var(e_n)\big(\frac{n\alpha}{B^2}\big) \leq B^2\frac{\alpha}{B^2} = \alpha. \end{aligned}$$

Taking complements gives $P[\pi(h) \in I_n] \geq 1-\alpha$, as claimed. □

If the chain is in stationarity, or at least reaches stationarity within n iterations, then the bias is zero, and we obtain:

Corollary 1: Let $n \in \mathbb{N}$ be a fixed time such that the chain is in stationarity after n steps. Then if $n\,Var(e_n) \leq B^2$ for some $B > 0$, then for any significance level $0 < \alpha < 1$, the interval

$$I_n := (e_n - n^{-1/2}\alpha^{-1/2}B,\; e_n + n^{-1/2}\alpha^{-1/2}B)$$

is a non-asymptotic conservative $1-\alpha$ confidence interval, i.e.

$$P[\pi(h) \in I_n] \;\geq\; 1-\alpha.$$

Proof. By the stationarity assumption, $|E(e_n) - \pi(h)| = 0$, so we can apply Theorem 2 with $C = 0$. It follows that $\delta = 0$ and $a_n = n^{-\frac{1}{2}}\alpha^{-1/2}B$. The result then follows immediately from Theorem 2. □

The assumptions for the non-asymptotic bound in Theorem 2 and Corollary 1 are stronger than for the asymptotic bound of Theorem 1, since they require a bound on the bias or for the chain to be at stationarity after n iterations. However, we will see in the next section that we can sometimes utilize properties such as polynomial ergodicity to help us establish a bound on bias. Also, in practice, MCMC users often approximately verify stationarity through a plethora of convergence diagnostics such as plots and renewal theory and non-parametric tests; see e.g. [12] for a review.

Next, we present a result which does not assume stationarity, nor require a bound on the bias, nor require a bound on the variance. But as a trade-off, it assumes a bound on an absolute first moment, which might be harder to verify. It could still be useful if e.g. the first moment condition can be linked to another property that is easier to satisfy, which could be explored in future research.

Theorem 3: Suppose for some fixed $n \in \mathbb{N}$ we have $E(|e_n - \pi(h)|) \leq \gamma_n$ for some constant $\gamma_n > 0$. Then for any significance level $\alpha \in (0, 1)$, the interval

$$I_n := (e_n - \gamma_n \alpha^{-1},\ e_n + \gamma_n \alpha^{-1})$$

is a non-asymptotic conservative $1 - \alpha$ confidence interval, i.e.

$$P[\pi(h) \in I_n] \ \geq \ 1 - \alpha.$$

Proof. Setting $a_n = \gamma_n/\alpha > 0$, we have by Markov's inequality that

$$\begin{aligned} P(|e_n - \pi(h)| \geq a_n) &\leq E(|e_n - \pi(h)|) \;/\; a_n \\ &= E(|e_n - \pi(h)|)\, \frac{\alpha}{\gamma_n} \\ &\leq (\gamma_n \frac{\alpha}{\gamma_n}) = \alpha. \end{aligned}$$

Taking complements,

$$\begin{aligned} P[\pi(h) \in I_n] &= P\big(|e_n - \pi(h)| \leq \gamma_n \alpha^{-1}\big) \\ &= P\big(|e_n - \pi(h)| \leq a_n\big) \geq 1 - \alpha. \end{aligned}$$ □

In particular, if the γ_n converge monotonically to zero, then we obtain a confidence interval which shrinks to a point as n approaches infinity.

4 The Order of MCMC Bias

Since we are estimating the quantity $\pi(h) = \int_{x \in \mathcal{X}} h(x)\, \pi(x)\, dx$ by the Markov chain estimator $e_n := \frac{1}{n} \sum_{i=1}^{n} h(X_i)$, the bias after n iterations is given by $\mathbf{Bias}(e_n) := E(e_n) - \pi(f)$. As previously mentioned, the results in [14] assumed this bias was $o(1/\sqrt{n})$ since it is generally believed to be $O(1/n)$ (see e.g. p. 21 of [4]). However, this is not always the case:

Example 1: Consider the Markov chain with state space $\mathcal{X} = \{0, 1, 2, 3, \dots\}$, and transition probabilities given by $p_{0,0} = 1$, and for all $n \geq 1$, $p_{n,0} = 1 - \frac{\sqrt{n}}{\sqrt{n+1}}$ and $p_{n,n+1} = \frac{\sqrt{n}}{\sqrt{n+1}}$. Then the chain is ϕ-irreducible (and aperiodic) with $\pi(x) = \phi(x) = \delta_0(x)$, i.e. $\pi(0) = 1$ and $\pi(x) = 0$ for all $x \neq 0$. Assume $X_0 = 1$. We then compute that, for $n = 1, 2, 3, \dots$,

$$P[X_n \neq 0] = P[X_n = n+1] = \prod_{i=1}^{n} \frac{\sqrt{i}}{\sqrt{i+1}} = \frac{1}{\sqrt{n+1}}.$$

Thus,

$$\lim_{n \to \infty} P[X_n \neq 0] = \lim_{n \to \infty} \frac{1}{\sqrt{n+1}} = 0.$$

So, the chain will converge to $\pi(\cdot)$ (from any initial distribution).

Next, consider the function on $\mathcal{X}$ defined by $f(0) = 0$, and $f(x) = 1$ for $x \geq 1$. Then $\pi(f) = f(0) = 0$. It follows that

$$\begin{aligned}
\mathbf{Bias}(e_n) &= E(e_n) - \pi(f) = E(e_n) = \frac{1}{n} \sum_{j=1}^{n} E[h(X_j)] \\
&= \frac{1}{n} \sum_{j=1}^{n} \Big[f(j+1) P(X_j = j+1) + f(0) P(X_j = 0) \Big] \\
&= \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{j+1}} \geq \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{n+1}} = \frac{1}{\sqrt{n+1}}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbf{Bias}(e_n) &= \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{j+1}} \leq \frac{1}{n} \int_{0}^{n} x^{-\frac{1}{2}}\, dx \\
&= \frac{1}{n}\, 2\, x^{\frac{1}{2}} \Big|_{x=0}^{x=n} = \frac{1}{n} (2\sqrt{n}) = \frac{2}{\sqrt{n}}.
\end{aligned}$$

That is, $\frac{1}{\sqrt{n+1}} \leq \mathbf{Bias}(e_n) \leq \frac{2}{\sqrt{n}}$. In particular, the bias is $O(1/\sqrt{n})$, but is not $O(1/n)$ nor even $o(1/\sqrt{n})$. $\square$

Example 1 raises the question of what conditions guarantee the bias to be $o(1/\sqrt{n})$. We shall derive such a result for a class of Markov chains that are polynomially ergodic, defined as follows (see e.g. [8], [9]):

Definition: Let $\{X_n\}$ be a Markov chain with stationary distribution $\pi(\cdot)$, and let $||\cdot||$ be total variation distance. Then the chain is *polynomially ergodic* if there exists a function $M : \mathcal{X} \to [0, \infty)$ such that:

$$||P^n(x, \cdot) - \pi(\cdot)|| \ \leq \ M(x)\, n^{-m}, \quad x \in \mathcal{X},\ x \in \mathbb{N};$$

here $m > 0$ is the *order* of the polynomial ergodicity.

Theorem 4: Let $\{X_n\}$ be a polynomially ergodic Markov chain of order $m > \frac{1}{2}$, with stationary distribution $\pi(\cdot)$. Suppose for some $D \in [0, \infty)$ and $f : \mathcal{X} \to \mathbb{R}$, we have $|f(x)| \leq D$. Then for any fixed initial state $X_0 = x$, the absolute bias $|\mathbf{Bias}(e_n)|$ is $o(n^{-1/2})$ as $n \to \infty$.

Proof. Let $X_0 = x$. We compute that

$$\begin{aligned}
|\mathbf{Bias}(e_n)| &= |E(e_n) - \pi(f)| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} |E[f(X_i)] - \pi(f)| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{g:\mathcal{X}\to\mathbf{R},\ |g(x)|\leq D} |E[g(X_i)] - \pi(g)| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} 2D\, ||P^i(x, \cdot) - \pi(\cdot)|| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} 2DM(x)\, i^{-m} \\
&= \frac{2DM(x)}{n} \sum_{i=1}^{n} i^{-m}.
\end{aligned}$$

Case 1: $\frac{1}{2} < m < 1$. Then

$$|\mathbf{Bias}(e_n)| \leq \frac{2DM(x)}{n} \sum_{i=1}^{n} i^{-m}$$

$$\leq \frac{2DM(x)}{n}\int_0^n x^{-m}dx$$
$$= \frac{2DM(x)}{n}\cdot\frac{1}{1-m}(n^{1-m}-0^{1-m})$$
$$= \frac{2DM(x)}{n}\cdot\frac{1}{1-m}(n^{1-m})$$

Therefore

$$n^{1/2}|\mathbf{Bias}(e_n)| \leq \frac{2DM(x)}{1-m}n^{1/2-m},$$

which $\to 0$ as $n \to \infty$ since $m > 1/2$ and $1-m > 0$ and D, $M(x) < \infty$.

Case 2: $m \geq 1$. Find some β such that $1/2 < \beta < 1 \leq m$. Then since $\sum_{i=1}^n i^{-m} \leq \sum_{i=1}^n i^{-\beta}$, it follows as above that

$$\lim_{n\to\infty} n^{1/2}|\mathbf{Bias}(e_n)| \leq \lim_{n\to\infty}\frac{2DM(x)}{1-\beta}n^{1/2-\beta} = 0. \qquad \square$$

Remark: This result says the bias is $o(1/\sqrt{n})$ for any polynomially ergodic chain of order more than 1/2. In the context of Theorem 2, this means that we can always find a constant C such that $|E(e_n) - \pi(h)| \leq C = \delta a_n$, since a_n is $O(1/\sqrt{n})$. Furthermore, if we can calculate an explicit value for $M(x)$, then we can obtain a value for C. As a specific example, if a chain has polynomial order $3/4 =: m$, with initial state $X_0 =: x$ satisfying $M(x) = 2$, and variance bound $n\,Var(e_n) \leq 4 =: B^2$, and functional bound $|f(x)| \leq 5 =: D$, then after $n = 100$ iterations we will have $|\mathbf{Bias}(e_n)| \leq \frac{2DM(x)}{n}\frac{1}{1-m}n^{1-m} = (20/n)(4)(n^{1/4}) \doteq 2.53 =: C$, so we can apply Theorem 2 at significance level $\alpha = 0.05$ to find that $\delta = \frac{2.53}{(2/\sqrt{5})+2.53} \doteq 0.74$ and $a_n = 2/(\sqrt{5}\,(1-0.74)) \doteq 3.44$, giving the 95% confidence interval $(e_n - 3.44,\, e_n + 3.44)$ after 100 iterations.

5 Extending to Non-Stationary Chains

Theorem 1 above was first proved assuming the chain started in stationarity. However, in practice MCMC is hardly ever started in stationarity, so accuracy bounds without this assumption are much more useful. We now prove a general result which says that asymptotic confidence intervals from stationarity can always be enlarged slightly to become asymptotic confidence intervals from arbitrary initial states.

Theorem 5: Consider an ergodic Markov chain $\{X_n\}$ on a state space $\mathcal{X}$ with stationary distribution $\pi(\cdot)$, functional h, and usual estimator e_n. Suppose the sequence $(e_n + a_n, e_n + b_n)$ is an asymptotic conservative $1-\alpha$ confidence interval for $\pi(h)$ when started in stationarity, i.e.

$$\liminf_{n\to\infty} P\big(a_n < \pi(h) - e_n < b_n\big) \;\geq\; 1-\alpha, \quad X_0 \sim \pi(\cdot).$$

Then for any $c > 0$ and $0 < r < 1$, and π-a.e. initial state $x \in \mathcal{X}$, the sequence $(e_n + a_n - cn^{-r}, e_n + b_n + cn^{-r})$ is an asymptotic conservative $1-\alpha$ confidence interval for the chain when started from the initial state $X_0 = x$, i.e.

$$\liminf_{n\to\infty} P\big(a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}\big) \;\geq\; 1-\alpha, \quad X_0 = x.$$

Proof. By ergodicity, for π-a.e. $x \in \mathcal{X}$, we have $\lim_{n\to\infty} ||P^n(x,\cdot) - \pi(\cdot)|| = 0$. Hence, for fixed $\epsilon > 0$ and $x \in \mathcal{X}$, we can find $m \in \mathbb{N}$ such that $||P^m(x,\cdot) - \pi(\cdot)|| \leq \epsilon$. Let $\{X_n\}$ be our original chain with $X_0 = x$, and let $\{X_n'\}$ be a second copy of the chain in stationarity, i.e. with $X_0' \sim \pi(\cdot)$ and hence $X_n' \sim \pi(\cdot)$ for all n. By Proposition 3(g) of [13], we can couple $\{X_n\}$ and $\{X_n'\}$ such that $P(H) \geq 1-\epsilon$, where

$$H \;=\; \{X_n = X_n' \text{ for all } n \geq m\}.$$

Now, let $e_n = \frac{1}{n}\sum_{i=1}^n h(X_i)$, and $e_n' = \frac{1}{n}\sum_{i=1}^n h(X_i')$ be the estimators from the two chains, so by assumption we have

$$\liminf_{n\to\infty} P\big(a_n < \pi(h) - e_n' < b_n\big) \;\geq\; 1-\alpha.$$

Then on the event H, for any $n \geq m$ we have

$$\Big|\big(\pi(h) - e_n\big) - \big(\pi(h) - e_n'\big)\Big| \;=\; \frac{1}{n}\Big|\sum_{i=1}^m \big(h(X_i') - h(X_i)\big)\Big| \;=:\; \frac{1}{n}|Z|,$$

where $Z = \big|\sum_{i=1}^m \big(h(X_i') - h(X_i)\big)\big|$. Hence, if H holds and $a_n < \pi(h) - e_n' < b_n$ and $\frac{1}{n}|Z| \leq cn^{-r}$, then $a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}$. Furthermore, Z is a fixed finite random variable, so there is $A < \infty$ with $P(|Z| > A) \leq \epsilon$. It follows that for $n \geq (A/c)^{1/(1-r)}$, we have

$$P\big(\frac{1}{n}|Z| > cn^{-r}\big) \;=\; P\big(|Z| > cn^{1-r}\big) \;\leq\; P\big(|Z| > A\big) \;\leq\; \epsilon.$$

We conclude that for all $n \geq \max[m,\ (A/c)^{1/(1-r)}]$,

$$\begin{aligned} & P\big(\{a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}\}^C\big) \\ \leq \ & P\big(\{a_n < \pi(h) - e'_n < b_n\}^C\big) \ + \ P(H^C) \ + \ P(|Z| > A) \\ \leq \ & P\big(\{a_n < \pi(h) - e'_n < b_n\}^C\big) \ + \epsilon + \epsilon, \end{aligned}$$

i.e.

$$P\big(a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}\big) \ \geq \ P\big(a_n < \pi(h) - e'_n < b_n\big) - 2\,\epsilon.$$

Then, taking $\liminf$ gives

$$\liminf_{n\to\infty} P\big(a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}\big) \ \geq \ \alpha - 2\,\epsilon.$$

Since this is true for any $\epsilon > 0$, we must actually have

$$\liminf_{n\to\infty} P\big(a_n - cn^{-r} < \pi(h) - e_n < b_n + cn^{-r}\big) \ \geq \ \alpha,$$

giving the result. □

6 Summary

In this paper, we have derived explicit asymptotic confidence intervals for any MCMC algorithm with finite asymptotic variance, started at any initial state, without requiring a Central Limit Theorem nor reversibility nor any bias bound. We have also derived explicit non-asymptotic confidence intervals assuming bounds on the bias or first moment, or alternatively that the chain starts in stationarity. We have related those non-asymptotic bounds to properties of MCMC bias, and shown that polynomially ergodicity implies appropriate bias bounds. It is our hope that these results will provide simple and useful tools for estimating errors of MCMC algorithms when CLTs are not available.

References

[1] M. Betancourt. Reply tweet of May 25, 2020. *Twitter*, 2020. Available at: https://twitter.com/betanalpha/status/1264960932255600647.

[2] S. Brooks, A. Gelman, G. Jones, and X. L. Meng. *Handbook of Markov chain Monte Carlo.* CRC press, 2011.

[3] C. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, 7:473–483, 1992.

[4] C. Geyer. Introduction to Markov chain Monte Carlo. *Handbook of Markov chain Monte Carlo*, 20116022:45, 2011.

[5] O. Häggström. On the central limit theorem for geometrically ergodic Markov chains. *Probability Theory and Related Fields*, 132(1):74–82, May 2005.

[6] O. Häggström and J.S. Rosenthal. On variance conditions for Markov chain CLTs. *Electronic Communications in Probability*, 12:454–464, 2007.

[7] I.A. Ibragimov and Y.V. Linnik. Independent and stationary sequences of random variables,(1971). *Walters-Noordhoff, Gröningen*, 1971.

[8] S.F. Jarner and R.L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli*, 9(4):559–578, 08 2003.

[9] G.L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

[10] C. Kipnis and S.R.S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[11] K. Latuszyński, B. Miasojedow, and W. Niemiro. Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli*, 19(5A):2033–2066, 2013.

[12] K.L. Mengersen, C.P. Robert, and C. Guihenneuc-Jouyaux. MCMC convergence diagnostics: a review. *Bayesian Statistics*, 6:415–440, 1999.

[13] G.O. Roberts and J.S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[14] J.S. Rosenthal. Simple confidence intervals for MCMC without CLTs. *Electron. J. Statist.*, 11(1):211–214, 2017.

[15] B. Tóth. Persistent random walks in random environment. *Probability Theory and Related Fields*, 71(4):615–625, Dec 1986.