MCMC Confidence Intervals and Biases Without CLTs

Jeffrey S. Rosenthal, University of Toronto
www.probability.ca / jeff@math.toronto.edu

(Algorithms Seminar, University of Warwick, January 22, 2021)

References:

J.S. Rosenthal, "Simple Confidence Intervals for MCMC Without CLTs". *Electronic Journal of Statistics* **11(1)** (2017), 211–214.

Y.H. Jiang, T. Liu, Z. Lou, J.S. Rosenthal, S. Shangguan, F. Wang, and Z. Wu, "MCMC Confidence Intervals and Biases". Submitted for publication & arXiv.org, December 2020.

Quotations:

"The CLT is the basis of all error estimation in Monte Carlo."
– Galin Jones, Computational Statistics (STAT 8701) Course Notes

"Moreover bias is of order n^{-1} ... whereas the standard error is of order $n^{-1/2}$, so bias is negligible in sufficiently long runs." – Charlie Geyer, Introduction to MCMC (Chapter 1 of the MCMC Handbook)

The Set-Up and Question:

Markov chain $\{X_n\}$ on $\mathcal{X}$, which converges to a stationary dist. $\pi(\cdot)$.

Functional $h : \mathcal{X} \to \mathbf{R}$, with finite expectation $\pi(h) := \mathbf{E}_\pi(h)$.

Estimate $\pi(h)$ by $e_n := \frac{1}{n}\sum_{i=1}^n h(X_i)$.

Want a "confidence" bound on how close e_n is to $\pi(h)$.

The Usual Approach – Use the CLT:

Under certain conditions, there is a $V > 0$, such that as $n \to \infty$,

$$e_n \;\approx\; N\Big(\pi(h),\; V/n\Big), \qquad \text{i.e.} \quad \frac{e_n - \pi(h)}{\sqrt{V/n}} \;\approx\; N(0,1)\,.$$

Hence, $\mathbf{P}\left(-1.96 \le \frac{e_n-\pi(h)}{\sqrt{V/n}} \le 1.96\right) \;\approx\; 0.95.$

So, as usual, $\mathbf{P}\left(\pi(h) \in \left[e_n - 1.96\sqrt{V/n},\ e_n + 1.96\sqrt{V/n}\right]\right) \approx 0.95.$

Gives a 95% confidence interval – good!

However, Markov chain CLTs only hold under certain conditions (which may be difficult to verify), such as:

* $\{X_n\}$ is uniformly ergodic, and $\mathbf{E}_\pi(|h|^2) < \infty$ (Cogburn 1972).

* $\{X_n\}$ is geometrically ergodic, and either (a) $\mathbf{E}_\pi(|h|^{2+\delta}) < \infty$ for some $\delta > 0$ (Ibragimov and Linnik, 1971), or (b) $\mathbf{E}_\pi(|h|^2) < \infty$ and the chain is reversible (Roberts and R., ECP 1997).

* $\{X_n\}$ is polynomially ergodic with suitable rate and moment bounds (Jarner and Roberts, AAP 2002).

But not always (Roberts, JAP 1999; Häggström, PTRF 2005).

What if we cannot establish a CLT? Assume it? Just give up?

Note: Can also get error bounds if we know the chain's conductance (Rudolf, J. Complexity 2009), or its regeneration tour structure (Latuszynski et al., Bernoulli 2013). But these may be unavailable.

Alternative Approach – Chebychev's Inequality (R., EJS 2017):

By the Triangle Inequality, for any $a_n > 0$,

$$\mathbf{P}\Big(|e_n - \pi(h)| \ge a_n\Big) = \mathbf{P}\Big(\Big|\big(e_n - \mathbf{E}(e_n)\big) + \big(\mathbf{E}(e_n) - \pi(h)\big)\Big| \ge a_n\Big)$$

$$\le \mathbf{P}\Big(|e_n - \mathbf{E}(e_n)| + |\mathbf{E}(e_n) - \pi(h)| \ge a_n\Big)$$

$$= \mathbf{P}\Big(|e_n - \mathbf{E}(e_n)| \ge a_n - |\mathbf{E}(e_n) - \pi(h)|\Big),$$

If $a_n - |\mathbf{E}(e_n) - \pi(h)| > 0$, then by Chebychev's Inequality this is

$$\le \mathbf{Var}(e_n) \Big/ \Big(a_n - |\mathbf{E}(e_n) - \pi(h)|\Big)^2.$$

Does this help? Well, assume that:

(A1) $V := \lim_{n\to\infty} n\,\mathbf{Var}(e_n) \in (0,\infty)$; [i.e. standard error is $O(1/\sqrt{n})$]

(A2) $\lim_{n\to\infty} n^{1/2}\,|\mathbf{E}(e_n) - \pi(h)| = 0$; [bias is $o(1/\sqrt{n})$, e.g. $O(1/n)$]

(A3) $\lim_{n\to\infty} \widehat{\sigma}_n^2 = V$. [i.e. have a consistent variance estimator]

And, let $\epsilon > 0$, and set $a_n = \sqrt{V(1+\epsilon)^2/n\alpha}$.

Then to first order as $n \to \infty$:

$a_n - |\mathbf{E}(e_n) - \pi(h)| \to a_n$ by (A2),

and $\mathbf{Var}(e_n) \to V/n$ by (A1),

which is $\le \widehat{\sigma}_n^2 (1+\epsilon)^2/n$ by (A3).

Hence,

$$\limsup_{n\to\infty} \mathbf{P}\Big(|e_n - \pi(h)| \ge a_n\Big) \;\le\; \limsup_{n\to\infty}\; \Big(\mathbf{Var}(e_n) \Big/ a_n^2\Big)$$

$$\le\; \limsup_{n\to\infty}\; \left(\widehat{\sigma}_n^2 (1+\epsilon)^2/n \Bigg/ \left[\sqrt{V(1+\epsilon)^2/n\alpha}\right]^2\right) \;=\; \alpha\,.$$

That is, $e_n \pm a_n$ is a conservative $(1-\alpha)$ confidence interval. (No CLT!)

Recall $a_n = \sqrt{V(1+\epsilon)^2/n\alpha}$. So, if $\alpha = 0.05$, and $\epsilon = 0.001$, then:

$$\mathbf{P}\left(\pi(h) \in [e_n - 4.48\sqrt{V/n},\; e_n + 4.48\sqrt{V/n}]\right) \;\gtrapprox\; 0.95.$$

Just like with CLT, except replace "1.96" by "4.48". (x 2.3)

Twitter Wars:

Simple! Easy! Everyone should use it! (But no one did ...)

Then, Michael Betancourt tweeted:

"The highest priority in any MCMC analysis is first verifying that a MCMC CLT holds." (Betancourt, 2019)

and "... under the nice conditions where a Markov chain Monte Carlo central limit theorem holds." (Betancourt, 2020)

My reply: "Or even without a central limit theorem! [link]"

His reply: "And how are A1 and A2 verified in practice?"

(A1) $V := \lim_{n\to\infty} n\, \mathbf{Var}(e_n) \in (0,\infty)$; (i.e. standard error is $O(1/n)$)

(A2) $\lim_{n\to\infty} n^{1/2}\, |\mathbf{E}(e_n) - \pi(h)| = 0$; (i.e. bias is $o(1/\sqrt{n})$)

Hm ... Is the MCMC bias always $O(1/n)$, or at least $o(1/\sqrt{n})$?

PROP: If $\{X_n\}$ is geometrically ergodic, or polynomially ergodic of order $> \frac{1}{2}$, then indeed $|\mathbf{E}(e_n) - \pi(h)|$ is $o(1/\sqrt{n})$.

But in general it might instead be $\Omega(1/\sqrt{n})$. (counter-example)

What to do?

Improved Result (Jiang et al., Preprint 2020):

THM: Assuming only that $\limsup_{n\to\infty} n\,\mathbf{Var}(e_n) \le V$, still have:

$$\limsup_{n\to\infty} \mathbf{P}\Big(|e_n - \pi(h)| \ge a_n\Big) \;\le\; \alpha$$

where again $\epsilon > 0$ and $a_n = \sqrt{V(1+\epsilon)^2/n\alpha}$.

(So, just need a variance bound, obtained from e.g. repeated runs, integrated autocorrelation times, batch means, window estimators, etc.)

PROOF: First assume $X_0 \sim \pi$: bias $= 0$, similar to above, $\epsilon = 0$.

Then generalise to π-a.e. $X_0 = x$, using a coupling.

Non-Asymptotic (!) Confidence Interval Version:

Suppose for a fixed $n \in \mathbf{N}$, $n\,\mathbf{Var}(e_n) \le V$, and $|\mathbf{E}(e_n) - \pi(h)| \le C$.

Then $\mathbf{P}\Big(|e_n - \pi(h)| \ge b_n\Big) \le \alpha$, where

$$b_n = \frac{\sqrt{V}}{\sqrt{n\alpha}\left(1 - \frac{C}{C+\sqrt{\frac{V}{n\alpha}}}\right)}.$$

(If $X_0 \sim \pi$, then $C = 0$, so $b_n = a_n$ from before, with $\epsilon = 0$. More generally, might hope that C would be small after sufficient burn-in.)

Seems useful to me! Use it!?!

All my papers, software, info: probability.ca/jeff