Fixed-Width Output Analysis for MCMC

Galin L. Jones
with James Flegal, Murali Haran, Ron Neath and Brian Caffo

School of Statistics
University of Minnesota

March 19, 2009

General Setting

Let π be a probability distribution. I want the value of some feature θ. For example, θ might be a quantile, a mode, an interval, or

$$\theta = E_\pi g := \int_{\mathsf{X}} g(x)\pi(dx)$$

Assume that θ is analytically intractable.

Treat θ as an unknown parameter and simulate data to estimate it.

Markov Chain Monte Carlo Basics

Simulate a Markov chain $X := \{X_n\}$

Use $\hat{\theta}_n = \hat{\theta}(X_0, X_1, \ldots, X_{n-1})$ to estimate θ so that

$$\hat{\theta}_n \to \theta \quad \text{as} \quad n \to \infty$$

<u>Usual Case</u>

$$\hat{\theta}_n = \bar{g}_n := \frac{1}{n} \sum_{i=0}^{n-1} g(X_i) \stackrel{a.s.}{\to} E_\pi g = \theta \ \text{ as } n \to \infty$$

Fixed-Width Methodology

When is n large enough?

When is $\hat{\theta}_n$ a good estimate of θ?

<u>Monte Carlo Error</u>: $\hat{\theta}_n - \theta$

<u>Sampling Distribution</u>

$$\tau_n(\hat{\theta}_n - \theta) \xrightarrow{d} J \ \text{ as } n \to \infty$$

Simulate until

$$[\hat{\theta}_n - c_n, \, \hat{\theta}_n + c_n]$$

is sufficiently narrow.

Fixed-Width Methodology

<u>Usual Case</u>

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \text{ as } n \to \infty$$

Simulate until

$$t_* \frac{\hat{\sigma}_g}{\sqrt{n}} + a(n) \leq \text{desired half-width}$$

where t_* is an appropriate critical value and $a(n) \downarrow 0$ on $\mathbb{Z}^+$.

Questions

Old Question

1. When is $\hat{\theta}_n$ a good estimate of θ?

New Questions

1. When does the Monte Carlo error have a limiting distribution?
2. How can we construct confidence intervals for θ?
3. Will the sequential procedure terminate at a finite time?
4. Will the resulting intervals have the desired coverage probability?

Regularity Conditions

$X = \{X_0, X_1, X_2, \ldots\}$ is a Markov chain

- invariant distribution is π
- π-irreducible
- aperiodic
- positive Harris recurrent

$$P^n(x, A) := \Pr(X_{i+n} \in A | X_i = x)$$

As $n \to \infty$

$$\|P^n(x, \cdot) - \pi(\cdot)\| := \sup_A |P^n(x, A) - \pi(A)| \downarrow 0$$

Regularity Conditions

Rate of TV convergence is the key:

$$\|P^n(x,\cdot) - \pi(\cdot)\| \leq C(x)t^n$$

where $C(x) \geq 0$ and $t \in (0,1)$.

Uniform / geometric ergodicity means C is bounded / unbounded.

There exist constructive techniques for establishing the rate of convergence.

Usual Case

$$\theta = E_\pi g$$

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \text{ as } n \to \infty$$

Simulate until

$$t_* \frac{\hat{\sigma}_g}{\sqrt{n}} + a(n) \leq \text{desired half-width}$$

where t_* is an appropriate critical value and $a(n) \downarrow 0$ on $\mathbb{Z}^+$.

Usual Case: CLT

Suppose at least one of the following conditions hold.

- X is uniformly ergodic and $E_\pi g^2 < \infty$
- X is geometrically ergodic and $E_\pi |g|^{2+\epsilon} < \infty$

Then for any initial distribution there exists $\sigma_g^2 \in (0, \infty)$ such that as $n \to \infty$

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2)$$

Usual Case: Estimating σ_g^2

Batch Means (nonoverlapping, overlapping, spaced)

Regenerative Simulation

Spectral Methods

Subsampling Bootstrap (overlapping batch means)

Time Series Bootstrap

Usual Case: Overlapping Batch Means

Split a long run $\{X_0, X_1, \ldots, X_{n-1}\}$ into batches of length a_n:

$$\begin{array}{ccc} X_0, \ldots, X_{a_n-1} & \qquad & \bar{g}_1 = \frac{1}{a_n}\sum_{j=0}^{a_n-1} g(X_j) \\ X_1, \ldots, X_{a_n} & & \bar{g}_2 \\ \vdots & & \vdots \end{array}$$

There are $n - a_n + 1$ batches of length a_n.

$$\hat{\sigma}^2_{OBM} = \frac{n a_n}{(n-a_n)(n-a_n+1)} \sum_{j=0}^{n-a_n} (\bar{g}_j - \bar{g}_n)^2$$

Usual Case: Overlapping Batch Means

Theorem

Suppose

- X is geometrically ergodic,
- $E_\pi |g(x)|^{2+\delta+\epsilon} < \infty$ for $\delta, \epsilon > 0$ and
- $a_n = \lfloor n^\nu \rfloor$ and $3/4 > \nu > (1 + \delta/2)^{-1}$,

then $\hat{\sigma}^2_{OBM} \to \sigma^2_g$ w.p. 1 as $n \to \infty$.

General Case

$\hat{\theta}_n$ approximates θ

Sampling Distribution

$$\tau_n(\hat{\theta}_n - \theta) \xrightarrow{d} J \;\text{ as } n \to \infty$$

Simulate until

$$[\hat{\theta}_n - c_n, \hat{\theta}_n + c_n]$$

is sufficiently narrow.

General Case: Subsampling Bootstrap

Split a long run $\{X_0, X_1, \dots, X_{n-1}\}$ into batches of length a_n:

$$\begin{array}{cc} X_0, \dots, X_{a_n-1} & \hat{\theta}_1 \\ X_1, \dots, X_{a_n} & \hat{\theta}_2 \\ \vdots & \vdots \end{array}$$

There are $n - a_n + 1$ batches of length a_n. The collection

$$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{n-a_n+1}$$

can be used to approximate the sampling distribution of $\hat{\theta}_n$.

Subsampling Bootstrap

<u>Theorem</u> Assume that as $n \to \infty$ $\tau_n \to \infty$ and

$$\tau_n(\hat{\theta}_n - \theta) \xrightarrow{d} J \, .$$

Let J^* be the empirical distribution function of the $\tau_{a_n}(\hat{\theta}_{a_n} - \hat{\theta}_n)$. If X is geometrically ergodic and as $n \to \infty$

1. $a_n \to \infty$ and $a_n/n \to 0$
2. $\tau_{a_n} \to \infty$ and $\tau_{a_n}/\tau_n \to 0$

then $J^* \to J$ at every continuity point and an "asymptotically valid" $100(1-\alpha)\%$ confidence interval for θ is

$$[\hat{\theta}_n - \tau_n^{-1} {J^*}^{-1}(1-\alpha/2), \, \hat{\theta}_n - \tau_n^{-1} {J^*}^{-1}(\alpha/2)] \, .$$

Baseball

Efron and Morris (1975) give a data set consisting of the raw batting averages (based on 45 official at-bats) and a transformation ($\sqrt{45}\arcsin(2x-1)$) for 18 Major League Baseball players during the 1970 season.

Suppose for $i = 1, \ldots, K$ that

$$\begin{aligned} Y_i|\gamma_i &\sim \mathrm{N}(\gamma_i, 1) & \gamma_i|\mu,\lambda &\sim \mathrm{N}(\mu,\lambda) \\ \lambda &\sim \mathrm{IG}(2,2) & f(\mu) &\propto 1\,. \end{aligned}$$

<u>Block Gibbs Sampler</u>: $(\lambda', \mu', \gamma') \to (\lambda, \mu, \gamma)$

<u>Theorem</u> (Rosenthal,1996) The Markov chain is geometrically ergodic.

Baseball

Goal: Estimate the posterior mean and median of γ_9, the "true" long-run (transformed) batting average of the Chicago Cubs' Ron Santo.

2000 Replications
Target half-width=.005
Nominal 95% confidence interval
Estimated Coverage Probability

OBM .948 (.003)
SS .951 (.005)

Summary

- Fixed-width methodology is useful in automating MCMC but requires a strongly consistent estimator of the asymptotic variance / asymptotically valid confidence interval.
- Fixed-width methods compare favorably to using diagnostics such as that developed by Gelman and Rubin.
- Spectral variance methods (Tukey-Hanning window) appear superior to batch means methods.
- The finite-sample properties of these methods have been extensively investigated and match the theory.
- There has been no assumption of stationarity.