Nonasymptotic bounds on the estimation error for regenerative MCMC[1]

Wojciech Niemiro

Nicolaus Copernicus University, Toruń and University of Warsaw
Poland

EPSRC Symposium Workshop on Markov Chain-Monte Carlo
Warwick, March 2009

[1] Work partially supported by Polish Ministry of Science and Higher Education Grant No. N N201387234.

Outline

Outline

Outline

1. Introduction
   - Computing integrals via MCMC
   - Accuracy bounds
2. Regeneration
   - Small set and regeneration
   - A sequential-regenerative estimator
3. Accuracy bounds
   - Mean Square Error
   - Confidence estimation and Median of Averages
   - How tight are the bounds?

Computing integrals via MCMC

We are to compute

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x =: \pi(f),$$

where

- $\mathcal{X}$ – state space,
- π – probability distribution on $\mathcal{X}$,

Markov chain

$$X_0, X_1, \ldots, X_t, \ldots \qquad \mathbb{P}(X_t \in \cdot) \to \pi(\cdot), \quad (t \to \infty).$$

MCMC estimator

$$\hat{\theta}_T = \frac{1}{T}\sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Computing integrals via MCMC

We are to compute

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x =: \pi(f),$$

where

- $\mathcal{X}$ – state space,
- π – probability distribution on $\mathcal{X}$,

Markov chain

$$X_0, X_1, \ldots, X_t, \ldots \qquad \mathbb{P}(X_t \in \cdot) \to \pi(\cdot), \quad (t \to \infty).$$

MCMC estimator

$$\hat{\theta}_T = \frac{1}{T}\sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Computing integrals via MCMC

We are to compute

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x =: \pi(f),$$

where

- $\mathcal{X}$ – state space,
- π – probability distribution on $\mathcal{X}$,

Markov chain

$$X_0, X_1, \ldots, X_t, \ldots \qquad \mathbb{P}(X_t \in \cdot) \to \pi(\cdot), \quad (t \to \infty).$$

MCMC estimator

$$\hat{\theta}_T = \frac{1}{T}\sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Accuracy bounds

Mean square error:

$$\text{MSE} \ = \ \mathbb{E}\,(\hat{\theta}_T - \theta)^2 \leq ?$$

Confidence bounds:

$$\mathbb{P}(|\hat{\theta}_T - \theta| > \varepsilon) \leq ?$$

Accuracy bounds

Mean square error:

$$\mathrm{MSE} = \mathbb{E}\,(\hat{\theta}_T - \theta)^2 \leq ?$$

Confidence bounds:

$$\mathbb{P}(|\hat{\theta}_T - \theta| > \varepsilon) \leq ?$$

Small set and regeneration

$$P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x).$$

ASSUMPTION (Small set)

There exist $J \subseteq \mathcal{X}$, a probability measure ν and $\beta > 0$ such that

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J) \nu(\cdot).$$

$Q(x, \cdot) := P(x, \cdot) - \beta \mathbb{I}(x \in J) \nu(\cdot)$ is the „residual" (sub-stochastic) kernel.

- If $X_{n-1} \notin J$ then draw $X_n \sim P(X_{n-1}, \cdot)$, no regeneration;
- If $X_{n-1} \in J$ then
  - with probability $1 - \beta$ draw $X_n \sim Q(X_{n-1}, \cdot)/(1 - \beta)$, no regeneration;
  - with probability β draw $X_n \sim \nu(\cdot)$, Regeneration.

Small set and regeneration

$$P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x).$$

ASSUMPTION (Small set)

There exist $J \subseteq \mathcal{X}$, a probability measure ν and $\beta > 0$ such that

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J) \nu(\cdot).$$

$Q(x, \cdot) := P(x, \cdot) - \beta \mathbb{I}(x \in J) \nu(\cdot)$ is the ,,residual" (sub-stochastic) kernel.

- If $X_{n-1} \notin J$ then draw $X_n \sim P(X_{n-1}, \cdot)$, no regeneration;
- If $X_{n-1} \in J$ then
  - with probability $1 - \beta$ draw $X_n \sim Q(X_{n-1}, \cdot)/(1-\beta)$, no regeneration;
  - with probability β draw $X_n \sim \nu(\cdot)$, Regeneration.

Small set and regeneration

$$P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x).$$

ASSUMPTION (Small set)

There exist $J \subseteq \mathcal{X}$, a probability measure ν and $\beta > 0$ such that

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J) \nu(\cdot).$$

$Q(x, \cdot) := P(x, \cdot) - \beta \mathbb{I}(x \in J) \nu(\cdot)$ is the ,,residual" (sub-stochastic) kernel.

- If $X_{n-1} \notin J$ then draw $X_n \sim P(X_{n-1}, \cdot)$, no regeneration;
- If $X_{n-1} \in J$ then
  - with probability $1 - \beta$ draw $X_n \sim Q(X_{n-1}, \cdot)/(1 - \beta)$, no regeneration;
  - with probability β draw $X_n \sim \nu(\cdot)$, Regeneration.

Regeneration

Remark: Actual sampling from $Q(X_{n-1}, \cdot)/(1-\beta)$ is **not** necessary (Mykland et al. 1995).

Start with $X_0 \sim \nu(\cdot)$ so that regeneration occurs at $T_0 = 0$.
Times of regeneration partition a trajectory into **iid blocks**:

$$\underbrace{X_0, \ldots, X_{T_1-1}}_{T_1}, \quad \underbrace{X_{T_1}, \ldots, X_{T_2-1}}_{T_2-T_1}, \quad \underbrace{X_{T_2}, \ldots, X_{T_3-1}}_{T_3-T_2}, \quad \cdots$$

$$\uparrow \qquad\qquad\qquad \uparrow \qquad\qquad\qquad \uparrow$$

R R R

R = Regeneration

Regeneration

Remark: Actual sampling from $Q(X_{n-1}, \cdot)/(1-\beta)$ is **not** necessary (Mykland et al. 1995).

Start with $X_0 \sim \nu(\cdot)$ so that regeneration occurs at $T_0 = 0$.
Times of regeneration partition a trajectory into **iid blocks**:

$$\underbrace{X_0, \ldots, X_{T_1-1}}_{T_1}, \quad \underbrace{X_{T_1}, \ldots, X_{T_2-1}}_{T_2-T_1}, \quad \underbrace{X_{T_2}, \ldots, X_{T_3-1}}_{T_3-T_2}, \quad \ldots$$

$\uparrow$ R $\quad \uparrow$ R $\quad \uparrow$ R

R = Regeneration

A sequential-regenerative estimator

$0 = T_0, T_1, \dots, T_r, \dots$ – moments of regeneration.

Fix n.

$$R(n) := \min\{r : T_r \geq n\}.$$

Estimator:

$$\hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i).$$

$$\begin{array}{ccccc} 0, \dots, T_1 - 1, & T_1, \dots\dots, & T_{R(n)-1}, \dots, & n, \dots, T_R(n) - 1, & T_{R(n)}, \dots \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathrm{R} & \mathrm{R} & \mathrm{R} & n & \mathrm{R} \end{array}$$

The estimator uses the part of trajectory written in blue

Mean Square Error

The integral of interest and its MCMC estimator:

$$\theta = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x, \qquad \hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i)$$

THEOREM

Under Assumption (Small set),

$$\begin{aligned} &(i) \quad MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma^2_{as}(f)}{n}\left(1 + \frac{\mu_2}{n}\right), \\ &(ii) \quad \mathbb{E}\,T_{R(n)} \leq n + \mu_2, \end{aligned}$$

where

$$\sigma^2_{as}(f) = \frac{1}{\mathbb{E}\,T_1}\mathbb{E}\left[\sum_{i=0}^{T_1-1}(f(X_i)-\theta)\right]^2, \qquad \mu_2 = \frac{\mathbb{E}\,T_1^2}{\mathbb{E}\,T_1} - 1.$$

Mean Square Error

The integral of interest and its MCMC estimator:

$$\theta = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x, \qquad \hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i)$$

THEOREM

Under Assumption (Small set),

$$(i) \quad MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma_{as}^2(f)}{n}\left(1 + \frac{\mu_2}{n}\right),$$

$$(ii) \quad \mathbb{E}\,T_{R(n)} \leq n + \mu_2,$$

where

$$\sigma_{as}^2(f) = \frac{1}{\mathbb{E}\,T_1}\mathbb{E}\left[\sum_{i=0}^{T_1-1}(f(X_i) - \theta)\right]^2, \qquad \mu_2 = \frac{\mathbb{E}\,T_1^2}{\mathbb{E}\,T_1} - 1.$$

Proof

(i) Two identities of Abraham Wald (sequential analysis):

If R is a stopping time then

$\mathbb{E}\sum_{i=1}^{R}\tau_k = \mathbb{E}R\mathbb{E}\tau_1$, where $\tau_k = T_k - T_{k-1}$,

$\mathrm{Var}\sum_{i=1}^{R} d_k = \mathbb{E}R\mathrm{Var}\, d_1$, where $d_k = \sum_{i=T_{k-1}}^{T_k-1}(f(X_i) - \theta)$.

(ii) Lorden's theorem (1970, renewal theory).

A geometric drift condition

ASSUMPTION (Drift)

There exist a function $u : \mathcal{X} \to [1, \infty)$*, constants* $\lambda < 1$ *and* $c < \infty$ *such that*

$$Qu(x) \leq \begin{cases} \lambda u(x) & \text{for } x \notin J, \\ c\lambda & \text{for } x \in J, \end{cases}$$

where $Q(x, \cdot) := P(x, \cdot) - \beta \mathbb{I}(x \in J)\nu(\cdot)$ *and* $Qu(x) := \int_{\mathcal{X}} Q(x, dy)u(y)$.

Explicit bounds under a drift condition

Bounds on $\sigma^2_{as}(f)$ and μ_2 which depend explicitly on the parameters λ, β, c and $\pi(u)$:

THEOREM

If Assumptions (Small set), (Drift) hold and $(f-\theta)^2 \le u$ *then*

$$\sigma^2_{as}(f) \le \pi(u)\frac{4A_1+9}{1-\lambda} + \sqrt{\pi(u)}\frac{A_2(1+A_1)}{(1-\lambda)^{3/2}},$$

$$\mu_2 \le 2\left(\pi(u)+1\right)\frac{1+A_1}{1-\lambda},$$

$$\textit{where } A_1 = \frac{\log c}{\log(1-\beta)^{-1}},$$

$$\textit{and } A_2 = 4\sqrt{\frac{c}{1-\beta}}.$$

Confidence estimation

Goal:

$$\mathbb{P}(|\hat{\theta} - \theta| \leq \varepsilon) \geq 1 - \alpha.$$

(given precision ε at a given level of confidence $1 - \alpha$).
We are to choose an estimator and a sufficient number of samples.

Possible approaches:

Chebyshev inequality:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\mathrm{MSE}}{\varepsilon^2}.$$

Problem: appears to be too loose.

Exponential inequalities? Problems with this approach:

- f is usually unbounded,
- involve quantities difficult to compute explicitly.

Confidence estimation

Goal:

$$\mathbb{P}(|\hat{\theta} - \theta| \leq \varepsilon) \geq 1 - \alpha.$$

(given precision ε at a given level of confidence $1 - \alpha$).
We are to choose an estimator and a sufficient number of samples.

Possible approaches:

Chebyshev inequality:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\text{MSE}}{\varepsilon^2}.$$

Problem: appears to be too loose.

Exponential inequalities? Problems with this approach:

- f is usually unbounded,
- involve quantities difficult to compute explicitly.

Confidence estimation

Goal:

$$\mathbb{P}(|\hat{\theta} - \theta| \leq \varepsilon) \geq 1 - \alpha.$$

(given precision ε at a given level of confidence $1 - \alpha$).
We are to choose an estimator and a sufficient number of samples.

Possible approaches:

Chebyshev inequality:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\text{MSE}}{\varepsilon^2}.$$

Problem: appears to be too loose.

Exponential inequalities? Problems with this approach:

- f is usually unbounded,
- involve quantities difficult to compute explicitly.

Median of Averages (MA)

Generate m independent copies of the Markov chain and compute estimators (averages):

$$
\begin{array}{lcl}
X_0^{(1)}, X_1^{(1)}, \ldots, X_t^{(1)}, \ldots & \longmapsto & \hat{\theta}^{(1)}, \\
\ldots & & \\
X_0^{(m)}, X_1^{(m)}, \ldots, X_t^{(m)}, \ldots & \longmapsto & \hat{\theta}^{(m)}.
\end{array}
$$

Estimator MA:

$$\hat{\theta} = \text{med}\left(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\right).$$

Median of Averages (MA)

$$\hat{\theta} = \text{med}\left(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\right).$$

Idea (Jerrum, Valiant and Vazirani, 1986):

Fix an initial moderate level of confidence $1 - a < 1 - \alpha$ and use Chebyshev inequality to get

$$\mathbb{P}(|\hat{\theta}^{(j)} - \theta| > \varepsilon) \leq a \qquad (j = 1, \ldots, m).$$

Then boost the level of confidence from $1 - a$ to $1 - \alpha$ by computing a median:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{1}{2}\exp\{-m/b\} = \alpha.$$

Optimizing the constants: $1 - a = 0.88031$, $b = 2.3147$
(Niemiro and Pokarowski, 2009, J. Appl. Probab.)

How tight are the bounds?

Asymptotic level of confidence, based on CLT:

$$\lim_{\varepsilon \to 0} \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) = \alpha,$$

for the number of samples

$$n \sim \frac{\sigma^2_{as}(f)}{\varepsilon^2} \left[\Phi^{-1}(1 - \alpha/2) \right]^2,$$

($\hat{\theta}$ is a simple average over n samples)

Nonasymptotic level of confidence:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \alpha,$$

for the **expected** number of samples

$$m\mathbb{E}T_{R(n)} \sim mn \sim C \cdot \frac{\sigma^2_{as}(f)}{\varepsilon^2} \log(2\alpha)^{-1},$$

($\hat{\theta}$ is MA with sequential/regenerative averages)

Symbol $\sim$ refers to $\alpha, \varepsilon \to 0$.

How tight are the bounds?

Asymptotic level of confidence, based on CLT:

$$\lim_{\varepsilon \to 0} \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) = \alpha,$$

for the number of samples

$$n \sim \frac{\sigma^2_{as}(f)}{\varepsilon^2} \left[\Phi^{-1}(1 - \alpha/2) \right]^2,$$

($\hat{\theta}$ is a simple average over n samples)

Nonasymptotic level of confidence:

$$\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \alpha,$$

for the **expected** number of samples

$$m \mathbb{E} T_{R(n)} \sim mn \sim C \cdot \frac{\sigma^2_{as}(f)}{\varepsilon^2} \log(2\alpha)^{-1},$$

($\hat{\theta}$ is MA with sequential/regenerative averages)

Symbol $\sim$ refers to $\alpha, \varepsilon \to 0$.

Complexity comparison

Asymptotic:

$$n \sim \frac{\sigma_{as}^2(f)}{\varepsilon^2}\left[\Phi^{-1}(1-\alpha/2)\right]^2,$$

Nonasymptotic:

$$m\mathbb{E}T_{R(n)} \sim mn \sim C \cdot \frac{\sigma_{as}^2(f)}{\varepsilon^2}\log(2\alpha)^{-1},$$

($\hat{\theta}$ is MA using sequential/regenerative averages)

$$\left[\Phi^{-1}(1-\alpha/2)\right]^2 \sim 2\log(2\alpha)^{-1}, \qquad (\alpha \to 0),$$

$$C \approx 19.34.$$

Complexity comparison

Asymptotic:

$$n \sim \frac{\sigma_{as}^2(f)}{\varepsilon^2}\left[\Phi^{-1}(1-\alpha/2)\right]^2,$$

Nonasymptotic:

$$m\mathbb{E}T_{R(n)} \sim mn \sim C \cdot \frac{\sigma_{as}^2(f)}{\varepsilon^2}\log(2\alpha)^{-1},$$

($\hat{\theta}$ is MA using sequential/regenerative averages)

$$\left[\Phi^{-1}(1-\alpha/2)\right]^2 \sim 2\log(2\alpha)^{-1}, \qquad (\alpha \to 0),$$

$$C \approx 19.34.$$

Problem

How to obtain practically applicable and explicitly computable bounds on $\sigma^2_{as}(f)$?