Explicit Error Bounds for Markov Chain Monte Carlo Methods

Daniel Rudolf

Mathematisches Institut
Friedrich Schiller Universität Jena

July 8, 2008
Outline

1. The problem
2. Error bounds
3. Application
4. Summary
The problem

Let \((\Omega, \mathcal{A}, \pi)\) be a probability space. Approximate

\[
S(f) = \mathbb{E}_\pi f = \int_\Omega f(x)\pi(dx).
\]

We assume that one cannot simulate \(\pi\) directly.

For instance let \(\varrho\) be an unnormalised density. Then

\[
S(f, \varrho) = \mathbb{E}_{\mu_\varrho} f = \frac{\int f(x)\varrho(x)dx}{\int \varrho(x)dx}.
\]

Method:
Simulate \(\pi\) with a Markov chain \(X_1, X_2, \ldots\) and compute

\[
S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}).
\]
The problem

Under known assumptions (ergodic theorem):

\[ S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}) \xrightarrow{n \to \infty} \int_{\Omega} f(x) \pi(dx) = S(f). \]

How does the error of \( S_{n,n_0} \) behave? Error bounds?

- **Error criterion:** \( e(S_{n,n_0}, f) = \left( \mathbf{E} |S(f) - S_{n,n_0}(f)|^2 \right)^{1/2} \).
- **Markov chain:** How to choose it?
- **Burn-in:** How large should \( n_0 \) be?
Properties of Markov chains

Let \((\Omega, \mathcal{A})\) be the state space and let \(K(\cdot, \cdot)\) be a Markov Kernel.

• distribution \(\pi\) is **stationary** if

\[
\pi(A) = \int_{\Omega} K(x, A) \pi(dx).
\]

• the Markov chain is **reversible** if

\[
\int_{A} K(x, B) \pi(dx) = \int_{B} K(x, A) \pi(dx).
\]

• the Markov chain is **lazy** if \(K(x, \{x\}) \geq 1/2\) for \(x \in \Omega\).

• the **conductance** is given as

\[
\varphi = \inf_{0 < \pi(A) \leq 1/2} \frac{\int_{A} K(x, A^c) \pi(dx)}{\pi(A)}.
\]
A known result

Theorem (Lovász and Simonovits (1993))

Let $X_1, X_2, \ldots$ be a lazy, reversible Markov chain with initial and stationary distribution $\pi$. Let $f \in L^2(\Omega, \pi)$ and $S_n(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_j)$. Then

$$e(S_n, f) \leq \frac{2}{\varphi \cdot \sqrt{n}} \|f\|_2.$$ 

- Laziness and reversibility are nonrestrictive conditions.
- The problem here is that the initial distribution is the stationary one.
Main result

Theorem (Rudolf (2007))

Let $X_1, X_2, \ldots$ be a lazy, reversible Markov chain. The initial distribution $\nu$ has a bounded density $\frac{d\nu}{d\pi}$ with respect to $\pi$. Then for $f \in L_\infty(\Omega, \pi)$ and $S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0})$ after a burn-in $n_0 \geq \frac{\log(\|d\nu/d\pi\|_\infty)}{\varphi^2}$ the error obeys $e(S_{n,n_0}, f) \leq \frac{10}{\varphi \cdot \sqrt{n}} \|f\|_\infty$.

- Burn-in time $n_0$ is explicitly given.
- To control the error and $n_0$ we need a lower bound for the conductance $\varphi$. 
**Application**

**Goal:** Let $\Omega \subset \mathbb{R}^d$. Approximate for given $f$ and $\varrho$

$$S(f, \varrho) = E_{\mu_\varrho} f = \frac{\int_\Omega f(x) \varrho(x) \, dx}{\int_\Omega \varrho(x) \, dx} \quad \text{with} \quad S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^n f(X_{j+n_0}).$$

- For densities with $\frac{\sup \varrho}{\inf \varrho} \leq C$ and $\|f\|_\infty \leq 1$ an almost optimal algorithm is known.
- If $C = 10^{20}$ and $n = 10^{17}$ we obtain for the worst case error $e_n > 0.7$.

So it is reasonable to consider a smaller class of functions.
Function class $\mathcal{F}^{\alpha}(\Omega)$

Definition of a class $\mathcal{F}^{\alpha}(\Omega)$ of functions with additional structure:

- $\varrho > 0$ is log-concave, for $x, y \in \Omega$ and $0 < \lambda < 1$
  \[ \varrho(\lambda x + (1 - \lambda)y) \geq \varrho(x)^{\lambda} \cdot \varrho(y)^{1-\lambda}. \]

- The logarithm of $\varrho$ is Lipschitz, i.e.,
  \[ |\log \varrho(x) - \log \varrho(y)| \leq \alpha \|x - y\|_2. \]
  This implies $\frac{\sup \varrho}{\inf \varrho} \leq e^{\alpha D}$ where $D$ is the diameter of $\Omega$.

- The functions $f$ are bounded such that $\|f\|_{\infty} \leq 1$. 
Metropolis with underlying ball walk

Let $\delta > 0$, let $\varrho$ be a positive density on $\Omega \subset \mathbb{R}^d$ and $B(x, \delta)$ be the ball with radius $\delta$ around $x$.

- choose $X_1$ randomly on $\Omega$;
- for $i = 1, \ldots, n + n_0$ do
  - if $\text{rand}() > 1/2$ then $X_{i+1} := X_i$;
  - else
    - choose $Y \in B(X_i, \delta)$ uniformly;
    - if $Y \notin \Omega$ then $X_{i+1} := X_i$;
    - if $Y \in \Omega$ and $\varrho(Y) \geq \varrho(X_i)$ then $X_{i+1} := Y$;
    - if $Y \in \Omega$ and $\varrho(Y) < \varrho(X_i)$ then
      - $X_{i+1} := Y$ with Prob $\varrho(Y)/\varrho(X_i)$ and
      - $X_{i+1} := X_i$ with Prob $1 - \varrho(Y)/\varrho(X_i)$.
- Return:

$$S_{n,n_0}^{\delta}(f, \varrho) = \frac{1}{n} \sum_{i=1}^{n} f(X_{i+n_0}).$$
A lower bound for the conductance

- Consider the \(d\)-dimensional unit ball, denoted by \(B^d\).
- This is important since the ball walk would get stuck with high probability in domains which have corners.

**Lemma (Mathé and Novak (2007))**

Let \(X_1, X_2, \ldots\) be the lazy Metropolis chain based on the ball walk, where \((f, \varphi) \in \mathcal{F}^\alpha(B^d)\). Then for \(\delta = \min \left\{ \frac{1}{\sqrt{d}+1}, \frac{1}{\alpha} \right\}\)

\[
\varphi \geq 0.00125 \frac{1}{\sqrt{d}+1} \min \left\{ \frac{1}{\sqrt{d}+1}, \frac{1}{\alpha} \right\}.
\]
Application of the theory

Corollary

Let \( X_1, X_2, \ldots \) be the lazy Metropolis chain which is based on a \( \delta \) ball walk, where \( \delta = \min \left\{ 1/\sqrt{d+1}, 1/\alpha \right\} \). Then for \( (f, \varrho) \in \mathcal{F}^\alpha(B^d) \)

\[
e(S_{n,n_0}^\delta, f) \leq 8000 \sqrt{d+1} \max \left\{ \sqrt{d+1}, \alpha \right\} \frac{\sqrt{d+1} \max \left\{ \sqrt{d+1}, \alpha \right\}}{\sqrt{n}},
\]

where

\[
n_0 \geq 1280000 \cdot \alpha (d+1) \max \left\{ d+1, \alpha^2 \right\}.
\]

- Hence the needed cost \( n + n_0 \) is proportional to

\[
d \max \left\{ d, \alpha^2 \right\} \cdot \alpha \varepsilon^{-2},
\]

i.e., it is polynomial in \( \alpha, d \) and \( \varepsilon^{-1} \).
Conclusion:

- An explicit error bound for Markov chain Monte Carlo integration is given, i.e. for

$$ n_0 \geq \frac{\log (\| \frac{d\nu}{d\pi} \|_\infty)}{\varphi^2} \quad \text{the error obeys} \quad e(S_n, n_0, f) \leq \frac{10}{\varphi} \cdot \sqrt{n} \| f \|_\infty. $$

- Hence the total cost $n_0 + n$ is bounded by

$$ n + n_0 \leq \frac{\log (\| \frac{d\nu}{d\pi} \|_\infty)}{\varphi^2} + \frac{100}{\varphi^2 \varepsilon^2} \| f \|_\infty^2. $$