Explicit error bounds for reversible Markov Chain Monte Carlo Methods

Daniel Rudolf
Mathematisches Institut
Friedrich Schiller Universität Jena

January 18, 2010

supported by
Outline

1 The problem
2 Error bounds
3 Examples
4 Summary
The problem

Let \((D, \pi)\) be a probability space. Approximate

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

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

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):

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

Error criterion:

\[ e_{\nu}(S_{n,n_0}, f) = \left( \mathbb{E}_{\nu,K} \left| S(f) - S_{n,n_0}(f) \right|^2 \right)^{1/2}. \]

(Markov chain with transition kernel \( K(\cdot, \cdot) \) and initial distribution \( \nu \))

- How does the error behave?
- Error bounds?
- How large should the burn-in \( n_0 \) be?
Properties of Markov chains

Let $D$ be the state space and let $K(\cdot, \cdot)$ be a Markov kernel.

- distribution $\pi$ is **stationary** if for all $A \subset D$ it holds
  \[ \pi(A) = \int_D K(x, A) \pi(dx). \]

- the Markov kernel is **reversible** if for all $A, B \subset D$ it holds
  \[ \int_A K(x, B) \pi(dx) = \int_B K(x, A) \pi(dx). \]

- the Markov chain has an $L_2$-**spectral gap** if
  \[ \beta := \| P - S \|_{L_2 \rightarrow L_2} < 1 \quad \text{where} \quad Pf(x) = \int_D f(y) K(x, dy). \]
Ergodicity

Let $D$ be the state space and let $K(\cdot, \cdot)$ be the Markov kernel.

- Let $\alpha \in [0, 1)$ and $M < \infty$ then the Markov kernel is called \((\alpha, M)\)-uniformly ergodic if
  \[
  \|K^n(x, \cdot) - \pi\|_{tv} \leq M\alpha^n, \quad \pi\text{-a.e.}, \quad n \in \mathbb{N}.
  \]

- The Markov kernel is called \textbf{geometric ergodic} if there exists $\alpha \in [0, 1)$ and $M(x) < \infty$ for all $x \in D$ such that
  \[
  \|K^n(x, \cdot) - \pi\|_{tv} \leq M(x)\alpha^n, \quad \pi\text{-a.e.}, \quad n \in \mathbb{N}.
  \]

- It is known: \textbf{$L_2$-spectral gap} $\iff$ \textbf{geometric ergodicity}.
Starting with stationary distribution

Theorem (similar to: Aldous 87, Lovász and Simonovits 93)

Let $X_1, X_2, \ldots$ be a reversible Markov chain with initial and stationary distribution $\pi$. Let

$$\Lambda = \sup \{ \lambda : \lambda \in \sigma(P - S) \} < 1.$$

Let $f \in L_2(\pi)$ and $S_n(f) = \frac{1}{n} \sum_{j=1}^n f(X_j)$. Then

$$\sup_{\|f\|_2 \leq 1} e_\pi(S_n, f)^2 = \frac{1 + \Lambda}{n(1 - \Lambda)} - \frac{2\Lambda(1 - \Lambda^n)}{n^2(1 - \Lambda)^2} \leq \frac{2}{n(1 - \Lambda)}.$$

- The initial distribution is the stationary one?
- Note that $\beta \geq \Lambda$. 

Daniel Rudolf (FSU Jena)  
Error bounds for MCMC Methods  
January 18, 2010  
7 / 29
Proposition

Let $X_1, X_2, \ldots$ be a reversible Markov chain with initial distribution $\nu$. Let $f \in L_2(\pi)$ and $g = f - S(f)$. Then

$$e_\nu(S_{n,n_0}, f)^2 = e_\pi(S_n, f)^2$$

$$+ \frac{1}{n^2} \sum_{j=1}^{n} L_{j+n_0}(g^2) + \frac{2}{n^2} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} L_{j+n_0}(g P^{k-j} g),$$

where

$$L_i(h) = \left\langle (P^i - S)(\frac{d\nu}{d\pi} - 1), h \right\rangle.$$
Starting with arbitrary distribution: $L_2$-result

Proposition

Let $X_1, X_2, \ldots$ be a reversible, $(\alpha, M)$-uniformly ergodic Markov chain. The initial distribution $\nu$ has a bounded density with respect to $\pi$.

Then for $f \in L_2(\pi)$ it holds

$$e_\nu(S_{n,n_0}, f)^2 \leq e_\pi(S_n, f)^2 + \frac{4M \| \frac{d\nu}{d\pi} - 1 \|_\infty \alpha^{n_0}}{n^2(1 - \alpha)^2} \| f \|_2^2.$$
Starting with arbitrary distribution: $L_p$-result

**Proposition**

Let $f \in L_p(\pi)$ with $p > 2$. Let $X_1, X_2, \ldots$ be a reversible, geometric ergodic Markov chain, i.e. $\beta < 1$.

- If $p \in (2, 4)$ then

$$e_\nu(S_{n_0}, f)^2 \leq e_\pi(S_n, f)^2 + \frac{16p^2}{(p - 2)} \frac{\beta^{2n_0(p-2)} p}{n^2(1 - \beta)^2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_p^p \left\| f \right\|_p^2.$$

- If $p \in [4, \infty]$ then

$$e_\nu(S_{n_0}, f)^2 \leq e_\pi(S_n, f)^2 + \frac{46\beta^{n_0} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2^2}{n^2(1 - \beta)^2} \left\| f \right\|_p^2.$$

Daniel Rudolf (FSU Jena)
Discussion

- Estimate has the right asymptotic order for \( n \to \infty \).
- If initial distribution \( \nu = \pi \) then set burn-in \( n_0 = 0 \).
- Dependence on \( p \)?

For \( p \in [4, \infty) \):

\[
\sup_{\|f\|_\rho \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta)} + \frac{46\beta^{n_0}}{n^2(1-\beta)^2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2^2.
\]

**Question:**

How to choose \( n_0 \) if the number of steps \( N = n_0 + n \) is given?
**Theorem:**

Let $p \geq 4$ and set burn-in

$$n_0 := \max \left\{ \left\lceil \log \left( \frac{\| d\nu \|_2 - 1}{1 - \beta} \right) \right\rfloor, 0 \right\}.$$ 

Then

$$\sup_{\| f \|_p \leq 1} e_\nu(S_{n_0}, n_0, f)^2 \leq \frac{2}{n(1 - \beta)} + \frac{46}{n^2(1 - \beta)^2}.$$
Burn-in

- Let $\beta = 0.99$ and $C := \| \frac{d\nu}{d\pi} - 1 \|_2 = 10^{30}$. 

\begin{align*}
N &= n_0 + n \\
n_0 &= \frac{\log(C)}{1 - \beta} \\
n_0 &= 0.88 \frac{\log(C)}{1 - \beta} \\
n_0 &= 3 \frac{\log(C)}{1 - \beta}
\end{align*}
Burn-in

- Let $\beta = 0.99$ and $C := \left\| \frac{d\nu}{d\pi} - 1 \right\|_2 = 10^{30}$. 

![Graph showing error bounds for MCMC Methods](graph.png)

- Error bounds for MCMC Methods
- $N = n_0 + n$
- $n_0 = \frac{\log(C)}{1-\beta}$
- $n_0 = 0.88 \frac{\log(C)}{1-\beta}$
- $n_0 = 3 \frac{\log(C)}{1-\beta}$
- $n_0 = 0$, init by $\pi$
Burn-in

- Let $\beta = 0.99$ and $C := \| \frac{d\nu}{d\pi} - 1 \|_2 = 10^{30}$. 

\begin{align*}
n_0 &= \frac{\log(C)}{1-\beta} \\
n_0 &= \frac{N}{2}, \ n = \frac{N}{2} \\
n_0 &= 0, \ \text{init by } \pi
\end{align*}
Literature

- Aldous 87: On the markov chain simulation method for uniform combinatorial distributions and simulated annealing.

- Lovász and Simonovits 93: Random Walks in a Convex Body and an Improved Volume Algorithm.


- Niemiro and Pokarowski 09: Fixed Precision MCMC Estimation by Median of Products of Averages.

- Łatuszyński and Niemiro 09: Rigorous confidence bounds for MCMC under a geometric drift condition.
Example: Integration over a convex body

Let $D \subset \mathbb{R}^d$. Approximate

$$S(f) = \int_D f(x) \mu(dx)$$

where $\mu$ is the uniform distribution on $D$.

- Let $D \subset \mathbb{R}^d$ be convex and $B(0, 1) \subset D \subset B(0, R)$, where $B(0, R)$ is the ball with radius $R$ around 0.

- Let $\ell$ be a line. Assume an oracle is available which gives the endpoints of

$$D \cap \ell.$$

- An oracle for $f$ is given.
Hit-and-run

- Hit-and-run Markov chain:

\[ D(0, 0) X_{1} r_{1} s_{1} D(0, 0) X_{1} r_{1} s_{1} X_{2} D(0, 0) X_{2} r_{2} s_{2} D(0, 0) X_{1} r_{2} s_{2} X_{3} D(0, 0) X_{1} r_{2} s_{2} X_{3} \]
Hit-and-run

Algorithm:

- choose $X_1$ uniformly in $B(0, 1)$.
- for $k = 1, \ldots, n + n_0$ do
  - if $\text{rand}() > 1/2$ then $X_{k+1} := X_k$;
  - else
    - choose uniformly a direction $d_k$ and determine $D \cap \{ X_k + \alpha d_k \mid \alpha \in \mathbb{R} \}$
    - with endpoints $r_k$ and $s_k$;
    - choose $X_{k+1}$ uniformly from $[r_k, s_k]$;
- return:
  $$S_{n,n_0}^\text{har} (f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i + n_0).$$
Properties of hit-and-run

- Hit-and-run is stationary and reversible with respect to $\mu$.

**Theorem (Lovász and Vempala 05)**

Let $D \subset \mathbb{R}^d$ be convex with $B(0, 1) \subset D \subset B(0, R)$. For the hit-and-run Markov chain it holds

$$1 - \beta \geq \frac{1}{2^{51}d^2R^2}.$$
Application of the theory

Corollary

Let $D \subset \mathbb{R}^d$ be convex with $B(0, 1) \subset D \subset B(0, R)$. Let $X_1, X_2, \ldots$ be the hit-and-run chain on $D$ and $p \geq 4$. Then

$$\sup_{\|f\|_p \leq 1} e(S_{n,n_0}^{\text{har}}, f) \leq 2^{26} \frac{d R}{\sqrt{n}} + 2^{29} \frac{d^2 R^2}{n},$$

where $n_0 \geq 2^{50} d^3 R^2 \log(R)$.

- Polynomial dependence on $d$ and $R$.
- Constants are very large. Practically not useful.
Example: Integration with an unknown density

Let $D \subset \mathbb{R}^d$. Approximate for $f$ and $\varrho$

$$S(f, \varrho) = E_{\mu_{\varrho}} f = \frac{\int_D f(x) \varrho(x) dx}{\int_D \varrho(x) dx}.$$

An oracle for $f$ and $\varrho > 0$ is available.

It is known:

- For densities with $\frac{\sup \varrho}{\inf \varrho} \leq C$ and $\|f\|_{\infty} \leq 1$ all algorithms are bad.
- If $C = 10^{20}$ and $n = 10^{17}$ we obtain for the worst case error $e_n > 0.7$.

One is forced to consider a smaller class of densities!
Function class $\mathcal{F}_\rho^\alpha(D)$

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

- $\varrho > 0$ is log-concave, for $x, y \in D$ 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 $\frac{\sup \varrho}{\inf \varrho} \leq e^{\alpha \cdot \text{dia}}$ where $\text{dia}$ is the diameter of $D$.

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

Let $\delta > 0$, let $(f, \varrho) \in \mathcal{F}_p^\alpha(D)$ and $B(x, \delta)$ be the ball with radius $\delta$ around $x$.

- choose $X_1$ uniformly in $D$;
- 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 $Y \notin D$ then $X_{i+1} := X_i$;
    - if $Y \in D$ and $\varrho(Y) \geq \varrho(X_i)$ then $X_{i+1} := Y$;
    - if $Y \in D$ 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 $1 - \beta$

- Consider the $d$-dimensional unit ball $B(0, 1)$.
- The ball walk would get stuck with high probability in domains which have corners.

**Lemma (Mathé and Novak 07)**

Let $X_1, X_2, \ldots$ be the lazy Metropolis chain based on the ball walk, where $(f, \varrho) \in \mathcal{F}^\alpha_p(B(0, 1))$. If $\delta = \min \left\{ \frac{1}{\sqrt{d} + 1}, \frac{1}{\alpha} \right\}$, then

$$1 - \beta \geq \frac{1}{1280000} \cdot \frac{1}{d + 1} \min \left\{ \frac{1}{d + 1}, \frac{1}{\alpha^2} \right\}.$$
Application of the theory

**Corollary**

Let $X_1, X_2, \ldots$ be the Metropolis chain which is based on a $\delta$ ball walk, where $\delta = \min \left\{ \frac{1}{\sqrt{d+1}}, \frac{1}{\alpha} \right\}$. Let $p \geq 4$. Then

$$\sup_{(f,\varrho) \in \mathcal{F}_p^\alpha (B(0,1))} e(S_{n,n_0}^\delta, f) \leq 1600 \frac{\sqrt{d+1} \max \left\{ \sqrt{d+1}, \alpha \right\}}{\sqrt{n}}$$

$$+ 8.7 \cdot 10^6 \frac{(d+1) \max \left\{ d+1, \alpha^2 \right\}}{n},$$

where

$$n_0 \geq 2.56 \cdot 10^6 \cdot \alpha (d+1) \max \left\{ d+1, \alpha^2 \right\}.$$
Discussion

- If \((f, \varrho) \in \mathcal{F}_\rho^\alpha(B(0,1))\) the needed cost \(n + n_0\) for solving the non-linear problem

\[
S(f, \varrho) = \frac{\int_{B(0,1)} f(x) \varrho(x) dx}{\int_{B(0,1)} \varrho(x) dx}
\]

within an error of \(\varepsilon\) is bounded by

\[
2.56 \cdot 10^6 \cdot (d + 1) \max \left\{ d + 1, \alpha^2 \right\} (\alpha + 4 \cdot \varepsilon^{-2}).
\]

- It is polynomial in \(\alpha\), \(d\) and \(\varepsilon^{-1}\), hence polynomial tractable.
Summary

- Markov chain uniformly ergodic
  \[ \rightarrow \] error bounds of \( S_{n,n_0} \) for \( f \in L_2(\pi) \).

- Markov chain geometric ergodic
  \[ \rightarrow \] error bounds of \( S_{n,n_0} \) for \( f \in L_p(\pi) \) with \( p > 2 \).

- Burn-in optimises the error bound.

- Tractability results

  (See also "Tractability of Multivariate Problems", work of Novak, Plaskota, Woźniakowski and others.)
Explicit error bound for Markov chain Monte Carlo:

If $p \geq 4$ and burn-in

$$n_0 := \max \left\{ \left\lceil \frac{\log \left( \| \frac{d\nu}{d\pi} - 1 \|_2 \right)}{1 - \beta} \right\rceil, 0 \right\}.$$ 

then

$$\sup_{\|f\|_p \leq 1} e_{\nu}(S_{n_0}, f)^2 \leq \frac{2}{n(1 - \beta)} + \frac{46}{n^2(1 - \beta)^2}.$$
Explicit error bound for Markov chain Monte Carlo:

If $p \geq 4$ and burn-in $n_0 := \max \left\{ \left\lceil \frac{\log \left( \frac{d\nu}{d\pi} - 1 \right)_2}{1 - \beta} \right\rceil, 0 \right\}$.

then

$$\sup_{\|f\|_p \leq 1} e_\nu(S_{n_0}, f)^2 \leq \frac{2}{n(1 - \beta)} + \frac{46}{n^2(1 - \beta)^2}.$$