Comparing three randomized algorithms for the hypercube

Mario Ullrich
Friedrich-Schiller-Universität Jena
Siegmundenburg, 18th July 2009
Overview

1 Introduction
   • Definitions
   • Algorithms

2 Variance computation
   • Basics

3 The hypercube
   • Definitions
   • Independence sampling
   • Importance sampling
   • Metropolis algorithm
   • Comparison

4 Literature
Problem

Let \((\mathcal{Z}, \mathcal{P}(\mathcal{Z}), \pi)\) be a finite probability space. For \(f : \mathcal{Z} \to \mathbb{R}\), we want to approximate

\[
S(f) = E_\pi(f) = \sum_{x \in \mathcal{Z}} f(x) \pi(x).
\]
Let \((\mathcal{Z}, \mathcal{P}(\mathcal{Z}), \pi)\) be a finite probability space. For \(f : \mathcal{Z} \rightarrow \mathbb{R}\), we want to approximate

\[ S(f) = E_{\pi}(f) = \sum_{x \in \mathcal{Z}} f(x) \pi(x). \]

The first idea is to generate random variables \(X_1, \ldots, X_n\) independent and identical distributed with respect to \(\pi\) to use the Simple Monte Carlo Method

\[ S_n^{\text{simple}}(f) = \frac{1}{n} \sum_{k=1}^{n} f(X_k). \]
Problem

Let \((\mathcal{Z}, \mathcal{P}(\mathcal{Z}), \pi)\) be a finite probability space. For \(f : \mathcal{Z} \to \mathbb{R}\), we want to approximate

\[
S(f) = E_{\pi}(f) = \sum_{x \in \mathcal{Z}} f(x) \pi(x).
\]

The first idea is to generate random variables \(X_1, ..., X_n\) independent and identical distributed with respect to \(\pi\) to use the **Simple Monte Carlo Method**

\[
S_n^{\text{simple}}(f) = \frac{1}{n} \sum_{k=1}^{n} f(X_k).
\]

But mostly we are not able to generate such variables. Therefore we have to run a Markov chain to generate variables that are approximately distributed w.r.t. \(\pi\). We will analyze only this case.
Let $P \in \mathbb{R}^{Z \times Z}$ be a transition (i.e. stochastic) matrix. Then

**Definition (reversible)**

$P$ is called *reversible* with respect to $\pi$ if the detailed balance condition holds true, i.e.

$$\pi(x) P(x, y) = \pi(y) P(y, x) \quad \forall x, y \in \mathbb{Z}.$$
Let $P \in \mathbb{R}^{\mathbb{Z} \times \mathbb{Z}}$ be a transition (i.e. stochastic) matrix. Then

**Definition (reversible)**

$P$ is called *reversible* with respect to $\pi$ if the detailed balance condition holds true, i.e.

$$\pi(x) P(x, y) = \pi(y) P(y, x) \quad \forall x, y \in \mathbb{Z}.$$ 

**Definition (ergodic)**

$P$ is called *ergodic* if $P$ is *irreducible*, i.e.

$$\forall x, y \in \mathbb{Z} \exists n > 0 : P^n(x, y) > 0$$

and $P$ is *aperiodic*, i.e.

$$\forall x \in \mathbb{Z} : \ GCD\{n > 0 : P^n(x, x) > 0\} = 1.$$
Assumption (1)

We have available a ergodic transition matrix $Q$ on $\mathcal{Z}$ which is reversible with respect to a probability measure $\sigma(x) > 0$ for all $x$ in $\mathcal{Z}$, so $\sigma$ is the unique stationary distribution of the Markov chain, i.e.

$$\sigma = \sigma \cdot Q.$$
Assumptions

Assumption (1)

We have available an ergodic transition matrix $Q$ on $\mathcal{Z}$ which is reversible with respect to a probability measure $\sigma(x) > 0$ for all $x$ in $\mathcal{Z}$, so $\sigma$ is the unique stationary distribution of the Markov chain, i.e.

$$\sigma = \sigma \cdot Q.$$ 

Assumption (2)

The initial distribution of the Markov chain is always the stationary one.  
This could be interpreted as an infinite burn-in of the Markov chain.
Independence sampling

We assume that the distribution $\sigma$ is simulatable. Let $X_1, \ldots, X_n$ be i.i.d. random variables with respect to $\sigma$. Then the \textit{Independence sampling estimator} is defined by

$$\tilde{S}_n^I(f) = \frac{1}{n} \sum_{k=1}^{n} \frac{\pi(X_k)}{\sigma(X_k)} f(X_k).$$
Algorithms

Independence sampling

We assume that the distribution $\sigma$ is simulatable. Let $X_1, ..., X_n$ be i.i.d. random variables with respect to $\sigma$. Then the Independence sampling estimator is defined by

$$\tilde{S}_n^I(f) = \frac{1}{n} \sum_{k=1}^{n} \frac{\pi(X_k)}{\sigma(X_k)} f(X_k).$$

Importance sampling

Let $X_1, ..., X_n$ be a realization of the Markov chain $(Q, \sigma)$. Then the Importance sampling estimator is defined by

$$S_n^I(f) = \frac{1}{n} \sum_{k=1}^{n} \frac{\pi(X_k)}{\sigma(X_k)} f(X_k).$$
**Metropolis algorithm**

We define

\[ M(x, y) = \begin{cases} 
Q(x, y) \alpha(x, y), & \text{if } x \neq y \\
Q(x, x) + \sum_{z \neq x} Q(x, z)(1 - \alpha(x, z)), & \text{if } x = y 
\end{cases} \]

where \( \alpha(x, y) = \min \left\{ 1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right\} \).

Let \( X_1, ..., X_n \) be a realization of the Markov chain \( (M, \pi) \), then the *Metropolis estimator* is defined by

\[ S_n^M(f) = \frac{1}{n} \sum_{k=1}^{n} f(X_k). \]
Variance computation

Let $P$ be a ergodic, reversible transition matrix with respect to $\pi$ and $L^2(\pi) = \{ f : Z \to \mathbb{R} \} = \mathbb{R}^Z$ with the scalar product

$$\langle f, g \rangle_\pi = E_\pi(f \ g) = \sum_{x \in Z} f(x) g(x) \pi(x).$$
Let $P$ be a ergodic, reversible transition matrix with respect to $\pi$ and $L^2(\pi) = \{f : \mathbb{Z} \to \mathbb{R}\} = \mathbb{R}^\mathbb{Z}$ with the scalar product

$$\langle f, g \rangle_\pi = E_\pi(f \ g) = \sum_{x \in \mathbb{Z}} f(x) g(x) \pi(x).$$

Let us consider $Pf(x) = \sum_{y \in \mathbb{Z}} P(x, y) f(y)$. Then reversibility of $P$ is equivalent to $P : L^2 \to L^2$ being self-adjoint. The spectral theorem implies that $P$ has real eigenvalues $1 = \beta_0 > \beta_1 \geq \beta_2 \geq \ldots \geq \beta_{|\mathbb{Z}|-1} > -1$ with an orthonormal basis of eigenfunctions $\psi_k : \mathbb{Z} \to \mathbb{R}$, so $P\psi_k = \beta_k \psi_k$ and $\langle \psi_k, \psi_j \rangle = \delta_{kj}$. 
Definitions

Definition (mean square error)

The mean square error of a MCMC-method $S_n^P$ is defined by

$$e(S_n^P, f) = \left( E_{\pi,P} | S_n^P(f) - S(f) |^2 \right)^{\frac{1}{2}},$$

where $E_{\pi,P}$ is the expectation with respect to $(P, \pi)$. 
Definitions

**Definition (mean square error)**

The *mean square error* of a MCMC-method $S_n^P$ is defined by

$$ e(S_n^P, f) = \left( \mathbb{E}_{\pi,P} |S_n^P(f) - S(f)|^2 \right)^{1/2}, $$

where $\mathbb{E}_{\pi,P}$ is the expectation with respect to $(P, \pi)$.

**Definition (asymptotic variance)**

The *asymptotic variance* is defined by

$$ \sigma^\infty(S_n^P, f) = \lim_{n \to \infty} n \cdot e(S_n^P, f)^2. $$
Proposition (Error of $S_n^{simple}$ and $\tilde{S}_n^I$)

Let $f \in L^2(\pi)$. Then

(i) $e(S_n^{simple}, f) = \frac{1}{\sqrt{n}} \text{Var}_\pi(f)$

(ii) $e(\tilde{S}_n^I, f) = \frac{1}{\sqrt{n}} \text{Var}_\sigma\left(\frac{\pi}{\sigma} f\right)$. 
Proposition (Error of $S_{n}^{\text{simple}}$ and $\tilde{S}_{n}$)

Let $f \in L^2(\pi)$. Then

$(i)$ $e(S_{n}^{\text{simple}}, f) = \frac{1}{\sqrt{n}} \text{Var}_{\pi}(f)$

$(ii)$ $e(\tilde{S}_{n}, f) = \frac{1}{\sqrt{n}} \text{Var}_{\sigma}\left(\frac{\pi}{\sigma} f\right)$.

Proposition (worst-case-error of $S_{n}^{\text{simple}}$ and $\tilde{S}_{n}$)

We get

$(i)$ $\sup_{\|f\|_2, \pi \leq 1} e(S_{n}^{\text{simple}}, f)^2 = \frac{1}{n}$

$(ii)$ $\frac{1}{n} \left(\left\|\frac{\pi}{\sigma}\right\|_{\infty} - 1\right) \leq \sup_{\|f\|_2, \pi \leq 1} e(\tilde{S}_{n}, f)^2 \leq \frac{1}{n} \left\|\frac{\pi}{\sigma}\right\|_{\infty}$.
Proposition (Bassetti und Diaconis)

Let \( f \in L^2(\pi) \) and \( X_1, \ldots, X_n \) be a realization of \((P, \pi)\). Then

\[
S^n_P(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i)
\]

has

\[
e(S^n_P, f)^2 = \frac{1}{n^2} \sum_{k=1}^{|Z|-1} a_k^2 W_n(\beta_k)
\]

where

\[
a_k = \langle f, \psi_k \rangle_\pi \quad \text{and} \quad W_n(\beta_k) = \frac{n - 2\beta_k - n\beta_k^2 + 2\beta_k^{n+1}}{(1 - \beta_k)^2}.
\]
Corollary (worst-case-error)

Applying the last proposition on \( \psi_1 \) it follows

\[
\sup_{\|f\|_{2, \pi} \leq 1} e(S_n^P, f)^2 = e(S_n^P, \psi_1)^2 = \frac{1 + \beta_1}{n(1 - \beta_1)} - \frac{2\beta_1(1 - \beta_1^n)}{(n(1 - \beta_1))^2}.
\]
Worst-case-error

**Corollary (worst-case-error)**

*Applying the last proposition on \( \psi_1 \) it follows*

\[
\sup_{\|f\|_{2,\pi} \leq 1} e(S_n^P, f)^2 = e(S_n^P, \psi_1)^2 = \frac{1 + \beta_1}{n(1 - \beta_1)} - \frac{2\beta_1(1 - \beta_1^n)}{(n(1 - \beta_1))^2}.
\]

**Corollary (asymptotic worst-case-variance)**

*In addition we get*

\[
\sigma_{\infty}^{\sup}(S_n^P) = \sup_{\|f\|_{2,\pi} \leq 1} \sigma_{\infty}(S_n^P, f) = \frac{1 + \beta_1}{1 - \beta_1}.
\]
Figure: worst-case-error for different $\beta_1$
Now we analyze the space $\mathbb{Z}_2^d = \{0, 1\}^d$ with the norm $|x| = \sum_{i=1}^{d} |x_i|$. 
Now we analyze the space $\mathbb{Z}_2^d = \{0, 1\}^d$ with the norm $|x| = \sum_{i=1}^{d} |x_i|$.

Furthermore

$$\pi(x) = \theta^{|x|} (1 - \theta)^{d-|x|}, \quad x \in \mathbb{Z}, \quad \text{where } \theta \in \left[\frac{1}{2}, 1\right],$$

and we have an ergodic transition matrix $Q$ with stationary distribution

$$\sigma(x) = p^{|x|} (1 - p)^{d-|x|}, \quad x \in \mathbb{Z}, \quad \text{where } p \in (0, 1)$$

and $p \neq \theta$. 

Mario Ullrich
Randomized algorithms
The used Markov chains on $\mathbb{Z}_2^d$ have the structure

$$P(x, y) = \frac{1}{d} \sum_{k=1}^{d} P^*(x_k, y_k) \prod_{j:j \neq k} 1\{x_j = y_j\}.$$ 

It is immediately seen, that every realization of $X_t$ has only $d$ neighbors.
The used Markov chains on $\mathbb{Z}_2^d$ have the structure

$$P(x, y) = \frac{1}{d} \sum_{k=1}^{d} P^*(x_k, y_k) \prod_{j: j \neq k} \mathbf{1}\{x_j = y_j\}.$$ 

It is immediately seen, that every realization of $X_t$ has only $d$ neighbors.

In our case the given transition matrix $Q$ holds

$$Q^* = \begin{pmatrix} \bar{p} & p \\ \bar{p} & p \end{pmatrix}, \quad \text{where} \quad \bar{p} = (1 - p).$$
Proposition (Independence sampling)

Let

\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\}. \]

Then by applying the estimate of the error of the Independence sampling, we get

\[ \kappa^d - 1 \leq \sigma^\infty(\tilde{S}_n^I) \leq \kappa^d. \]
Proposition (Independence sampling)

Let

\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\}. \]

Then by applying the estimate of the error of the Independence sampling, we get

\[ \kappa^d - 1 \leq \sigma_{\infty}^\sup(\tilde{S}_n^l) \leq \kappa^d. \]

\[ \longrightarrow \sigma_{\infty}^\sup(\tilde{S}_n^l) \text{ increases exponentially in } d. \]
Importance sampling

We need the eigenvalues and -functions of $Q$ with $Q^* = \begin{pmatrix} \bar{p} & p \\ \bar{p} & p \end{pmatrix}$. 
Importance sampling

We need the eigenvalues and -functions of $Q$ with $Q^* = \begin{pmatrix} \bar{p} & p \\ \bar{p} & p \end{pmatrix}$.

These are

$$\beta_z = 1 - \frac{|z|}{d}$$

and

$$\psi_z(x) = \prod_{i=1}^{d} \psi_{z_i}(x_i) = \prod_{i=1}^{d} \left( \sqrt{\frac{p}{\bar{p}}} \right)^{z_i} (1-x_i) \left( -\sqrt{\frac{p}{\bar{p}}} \right)^{z_i x_i}$$

which are indexed by $z = (z_1, ..., z_d) \in \mathbb{Z}_2^d$.

These functions are orthonormal in $L^2(\sigma)$. 
Importance sampling

We need the eigenvalues and -functions of $Q$ with $Q^* = \begin{pmatrix} \bar{p} & p \\ \bar{p} & p \end{pmatrix}$.

These are

$$\beta_z = 1 - \frac{|z|}{d}$$

and

$$\psi_z(x) = \prod_{i=1}^{d} \psi_{z_i}(x_i) = \prod_{i=1}^{d} \left( \sqrt{\frac{p}{\bar{p}}} \right)^{z_i(1-x_i)} \left( - \sqrt{\frac{p}{\bar{p}}} \right)^{z_i x_i}$$

which are indexed by $z = (z_1, \ldots, z_d) \in \mathbb{Z}_2^d$.

These functions are orthonormal in $L^2(\sigma)$.

For the variance computation we need the fourier coefficients $a_z = \langle \frac{\pi}{\sigma} f, \psi_z \rangle_\sigma$. 
Proposition (Importance sampling)

For $\bar{\psi}_1 = \left(\sqrt{\frac{p}{\bar{p}}}\right)^{1-x_1} \left(-\sqrt{\frac{p}{\bar{p}}}\right)^{x_1} / \left(\bar{\theta} \frac{p}{\bar{p}} + \theta \frac{\bar{p}}{p}\right)^{1/2}$ with a suitable constant $K(\theta, p)$, that only depends on $\theta$ and $p$, we have

$$\sigma_\infty(S_n, \bar{\psi}_1) \sim K(\theta, p) \left(1 + \left(\bar{\theta} \sqrt{\frac{p}{\bar{p}}} - \theta \sqrt{\frac{\bar{p}}{p}}\right)^2\right)^d + O(d).$$
Proposition (Importance sampling)

For $\psi_1 = \left(\sqrt{\frac{p}{\bar{p}}} \right)^{(1-x_1)} \left( -\sqrt{\frac{p}{\bar{p}}} \right)^{x_1} / \left( \bar{\theta} \frac{p}{\bar{p}} + \theta \frac{\bar{p}}{p} \right)^{1/2}$ with a suitable constant $K(\theta, p)$, that only depends on $\theta$ and $p$, we have

$$\sigma_\infty(S_n^l, \bar{\psi}_1) \sim K(\theta, p) \left( 1 + \left( \bar{\theta} \sqrt{\frac{p}{\bar{p}}} - \theta \sqrt{\frac{\bar{p}}{p}} \right)^2 \right)^d + O(d).$$

$\longrightarrow \sigma_\infty^{sup}(S_n^l) \geq \sigma_\infty(S_n^l, \bar{\psi}_1)$ increases exponentially in $d$, too.
Metropolis algorithm

By Applying the Metropolis-Hastings-algorithm we get for $p \leq \theta$

$$M^* = \begin{pmatrix} \bar{p} & p \\ p \bar{\theta}/\theta & 1 - p \bar{\theta}/\theta \end{pmatrix}, \quad \text{mit } \bar{\theta} = (1 - \theta)$$

and for $\theta \leq p$

$$M^* = \begin{pmatrix} 1 - \theta \bar{p}/\bar{\theta} & \theta \bar{p}/\bar{\theta} \\ \bar{p} & p \end{pmatrix}, \quad \text{mit } \bar{\theta} = (1 - \theta).$$
Metropolis algorithm

By Applying the Metropolis-Hastings-algorithm we get for $p \leq \theta$

$$M^* = \begin{pmatrix} \bar{p} & p \\ \bar{p} & 1-p \bar{\theta}/\theta \end{pmatrix} , \quad \text{mit } \bar{\theta} = (1-\theta)$$

and for $\theta \leq p$

$$M^* = \begin{pmatrix} 1-\theta \bar{p}/\bar{\theta} & \theta \bar{p}/\bar{\theta} \\ \bar{p} & p \end{pmatrix} , \quad \text{mit } \bar{\theta} = (1-\theta).$$

With $\kappa = \max \left\{ \frac{\theta}{p}, \frac{1-\theta}{1-p} \right\}$ we have

$$\beta^*_z = 1 - \frac{|z|}{d \kappa}$$

$$\psi^*_z(x) = \prod_{i=1}^d \psi_{z_i}(x_i) = \prod_{i=1}^d \left( \sqrt{\frac{\theta}{\bar{\theta}}} \right)^{z_i(1-x_i)} \left( -\sqrt{\frac{\bar{\theta}}{\theta}} \right)^{z_i x_i}$$
Proposition (Metropolis algorithm)

Let
\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\}. \]

Then
\[ \sigma^\sup \left( S_n^M \right) = \sigma^\infty \left( S_n^M, \psi^*_1 \right) = 2d \kappa - 1. \]
Proposition (Metropolis algorithm)

Let

\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\}. \]

Then

\[ \sigma^\sup_\infty (S_n^M) = \sigma_\infty (S_n^M, \psi_1^*) = 2d \kappa - 1. \]

\[ \rightarrow \sigma^\sup_\infty (S_n^M) \text{ increases only linear in } d. \]
Comparison

Proposition

On $\mathbb{Z}^d_2$ with $d > 1$

$$\sigma_{\infty}^{\sup}(S_{n}^M) \leq \sigma_{\infty}^{\sup}(S_{n}^I),$$

holds true, if

$$p \notin I := \left( (2d)^{\frac{1}{1-d}} \theta, 1 - (2d)^{\frac{1}{1-d}} (1 - \theta) \right).$$

Moreover

$$I \xrightarrow{d \to \infty} \{\theta\}.$$
**Proposition**

*On $\mathbb{Z}_2^d$ with $d > 1* 

$$
\sigma_\infty(S^M_n) \leq \sigma_\infty(S^I_n),
$$

holds true, if

$$
p \notin I := \left( (2d)^{\frac{1}{1-d}} \theta, 1 - (2d)^{\frac{1}{1-d}} (1 - \theta) \right).
$$

Moreover

$$
I \xrightarrow{d \to \infty} \{\theta\}.
$$

→ $S^M_n$ is asymptotically (in $d$) better than $S^I_n$ while $p \neq \theta$. 
Proposition

Let

\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\} \]

and

\[ d \geq \frac{\ln 2/3}{\ln \theta}. \]

Then

\[ \frac{\sigma^\text{sup}_\infty (S^I_n)}{\sigma^\text{sup}_\infty (S^M_n)} \geq \frac{\kappa^{d-1}}{2d}. \]
Comparison

Proposition

Let

\[ \kappa = \max \left\{ \frac{\theta}{p}, \frac{1 - \theta}{1 - p} \right\} \]

and

\[ d \geq \frac{\ln 2/3}{\ln \theta}. \]

Then

\[ \frac{\sigma_{\infty}^\sup (S_n^I)}{\sigma_{\infty}^\sup (S_n^M)} \geq \frac{\kappa^{d-1}}{2d}. \]

\[ \rightarrow S_n^I \text{ is exponentially worse than } S_n^M. \]
Figure: asymptotic worst-case-variance for different $p$ and $\theta$
Thank you!
Literature

- D. Rudolf. Error bounds for computing the expectation by MCMC, *manuscript*