Randomized Quasi-Monte Carlo:

Theory, Choice of Discrepancy, and Applications

featuring randomly-shifted lattice rules

Pierre L'Ecuyer and David Munger

Informatique et Recherche Opérationnelle, Université de Montréal

Basic Monte Carlo setting

Want to estimate

$$\mu = \mu(f) = \int_{[0,1)^s} f(\mathbf{u})\, d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$$

where $f : [0,1)^s \to \mathbb{R}$ and $\mathbf{U}$ is a uniform r.v. over $[0,1)^s$.

Standard Monte Carlo:

- Generate n independent copies of $\mathbf{U}$, say $\mathbf{U}_1, \ldots, \mathbf{U}_n$;
- estimate μ by $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{U}_i)$.

Randomized quasi-Monte Carlo (RQMC)

An RQMC estimator of μ has the form

$$\hat{\mu}_{n,\mathrm{rqmc}} = \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{U}_i),$$

with $P_n = \{\mathbf{U}_0, \ldots, \mathbf{U}_{n-1}\} \subset (0,1)^s$ an RQMC point set:

(i) each point $\mathbf{U}_i$ has the uniform distribution over $(0,1)^s$;

(ii) P_n as a whole is a low-discrepancy point set.

$$\mathbb{E}[\hat{\mu}_{n,\mathrm{rqmc}}] = \mu \qquad \text{(unbiased)}.$$

Can perform m independent realizations $X_1, \ldots, X_m$ of $\hat{\mu}_{n,\mathrm{rqmc}}$, then estimate μ and $\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}]$ by their sample mean $\bar{X}_m$ and sample variance S_m^2 (also unbiased).

Temptation: assume that $\bar{X}_m$ has the normal distribution.

Generalized antithetic variates and RQMC

$$
\begin{aligned}
\mathrm{Var}[\hat\mu_{n,\mathrm{rqmc}}] &= \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\mathrm{Cov}[f(\mathbf{U}_i), f(\mathbf{U}_j)] \\
&= \frac{\mathrm{Var}[f(\mathbf{U}_i)]}{n} + \frac{2}{n^2}\sum_{i<j}\mathrm{Cov}[f(\mathbf{U}_i), f(\mathbf{U}_j)].
\end{aligned}
$$

We want to make the last sum as negative as possible.

Special cases:

antithetic variates ($n = 2$),
Latin hypercube sampling (LHS),
randomized quasi-Monte Carlo (RQMC).

Lattice rules

Integration lattice:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z} \right\},$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_s \in \mathbb{R}^s$ are linearly independent over $\mathbb{R}$ and where L_s contains $\mathbb{Z}^s$. Lattice rule: Take $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} = L_s \cap [0,1)^s$.

Lattice rules

Integration lattice:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z} \right\},$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_s \in \mathbb{R}^s$ are linearly independent over $\mathbb{R}$ and where L_s contains $\mathbb{Z}^s$. Lattice rule: Take $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} = L_s \cap [0,1)^s$.

Lattice rule of rank 1: $\mathbf{u}_i = i\mathbf{v}_1 \bmod 1$ for $i = 0, \ldots, n-1$,
where $n\mathbf{v}_1 = \mathbf{z} = (z_1, \ldots, z_s) \in \{0, 1, \ldots, n-1\}$.
Korobov rule: $\mathbf{z} = (1, a, a^2 \mod n, \ldots)$.

Lattice rules

Integration lattice:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z} \right\},$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_s \in \mathbb{R}^s$ are linearly independent over $\mathbb{R}$ and where L_s contains $\mathbb{Z}^s$. Lattice rule: Take $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} = L_s \cap [0,1)^s$.

Lattice rule of rank 1: $\mathbf{u}_i = i\mathbf{v}_1 \bmod 1$ for $i = 0, \ldots, n-1$,
where $n\mathbf{v}_1 = \mathbf{z} = (z_1, \ldots, z_s) \in \{0, 1, \ldots, n-1\}$.
Korobov rule: $\mathbf{z} = (1, a, a^2 \mod n, \ldots)$.

For any $\mathfrak{u} \subset \{1, \ldots, s\}$, the projection $L_s(\mathfrak{u})$ of L_s is also a lattice.

Lattice rules

Integration lattice:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z} \right\},$$

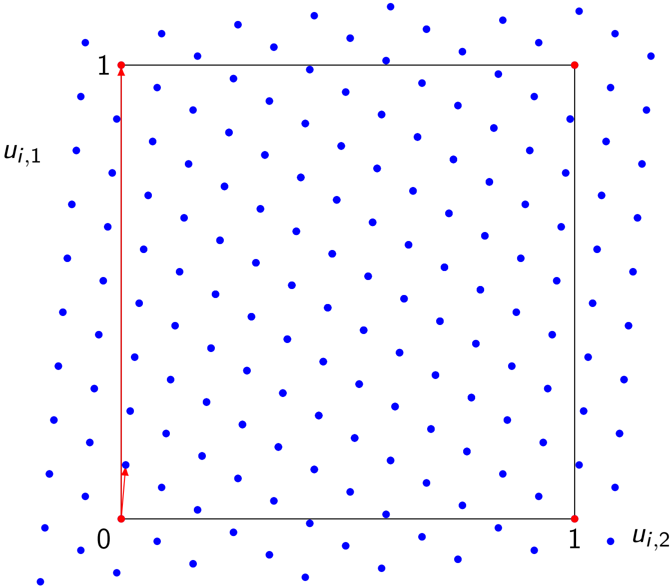
where $\mathbf{v}_1, \ldots, \mathbf{v}_s \in \mathbb{R}^s$ are linearly independent over $\mathbb{R}$ and where L_s contains $\mathbb{Z}^s$. Lattice rule: Take $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} = L_s \cap [0,1)^s$.

Lattice rule of rank 1: $\mathbf{u}_i = i\mathbf{v}_1 \bmod 1$ for $i = 0, \ldots, n-1$,
where $n\mathbf{v}_1 = \mathbf{z} = (z_1, \ldots, z_s) \in \{0, 1, \ldots, n-1\}$.
Korobov rule: $\mathbf{z} = (1, a, a^2 \mod n, \ldots)$.

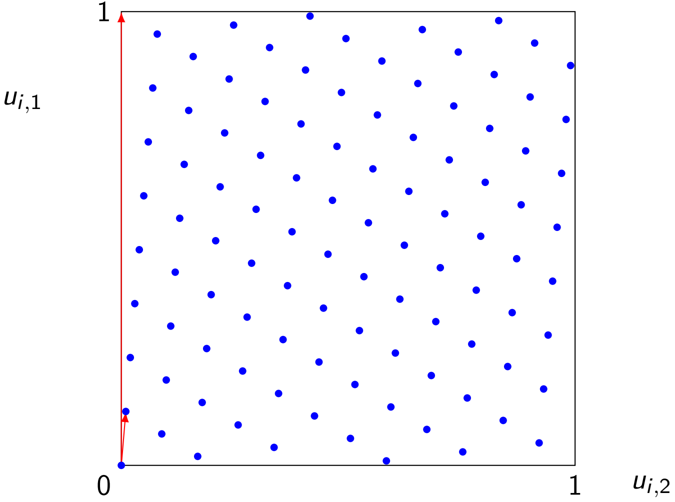
For any $\mathfrak{u} \subset \{1, \ldots, s\}$, the projection $L_s(\mathfrak{u})$ of L_s is also a lattice.

Random shift modulo 1: generate a single point $\mathbf{U}$ uniformly over $(0,1)^s$ and add it to each point of P_n, modulo 1, coordinate-wise:
$\mathbf{U}_i = (\mathbf{u}_i + \mathbf{U}) \bmod 1$. Each $\mathbf{U}_i$ is uniformly distributed over $[0,1)^s$.

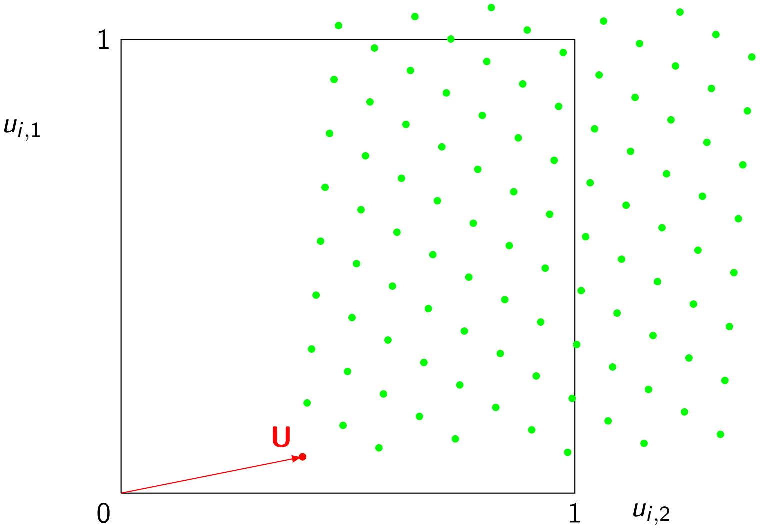
Randomly-shifted lattice: Two-dim. example



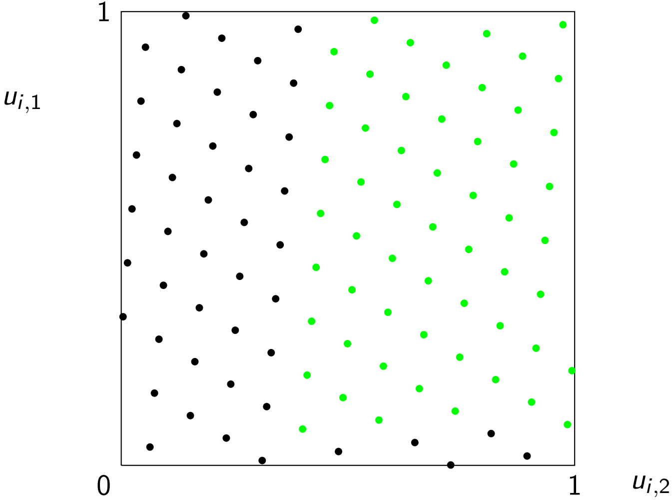
Randomly-shifted lattice: Two-dim. example



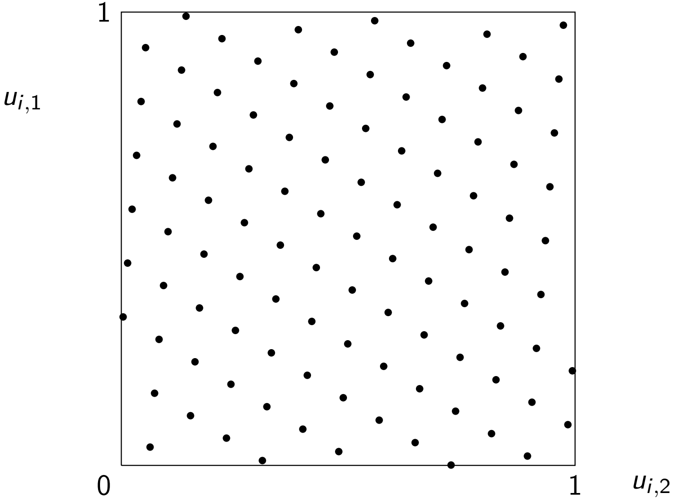
Randomly-shifted lattice: Two-dim. example



Randomly-shifted lattice: Two-dim. example



Randomly-shifted lattice: Two-dim. example



Variance expression

Suppose f has Fourier expansion

$$f(\mathbf{u}) = \sum_{\mathbf{h}\in\mathbb{Z}^s} \hat{f}(\mathbf{h}) e^{2\pi\sqrt{-1}\mathbf{h}^{\mathsf{t}}\mathbf{u}}.$$

For a randomly shifted lattice, the exact variance is (always)

$$\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} |\hat{f}(\mathbf{h})|^2,$$

where $L_s^* = \{\mathbf{h} \in \mathbb{R}^s : \mathbf{h}^{\mathsf{t}}\mathbf{v} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_s\} \subseteq \mathbb{Z}^s$ is the dual lattice.

From the viewpoint of variance reduction, an optimal lattice for f is one that minimizes $\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}]$.

$$\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} |\hat{f}(\mathbf{h})|^2.$$

If f has square-integrable mixed partial derivatives up to order α, and the periodic continuation of its derivatives up to order $\alpha-2$ is continuous across the unit cube boundaries, then

$$|\hat{f}(\mathbf{h})|^2 = \mathcal{O}((\max(1,h_1),\ldots,\max(1,h_s))^{-2\alpha}).$$

Moreover, there is a $\mathbf{v}_1 = \mathbf{v}_1(n)$ such that

$$\mathcal{P}_{2\alpha} \stackrel{\mathrm{def}}{=} \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} (\max(1,h_1),\ldots,\max(1,h_s))^{-2\alpha} = \mathcal{O}(n^{-2\alpha+\delta}).$$

This is the variance for a worst-case f having

$$|\hat{f}(\mathbf{h})|^2 = (\max(1,h_1),\ldots,\max(1,h_s))^{-2\alpha}.$$

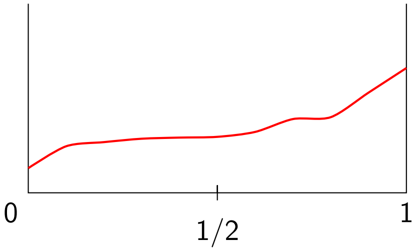
Thus, for a f that satisfies the above conditions, there are rank-1 lattices for which $\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \mathcal{O}(n^{-2\alpha+\delta})$.

Warning: the hidden constant in $\mathcal{O}$ can be large when s is large.

Baker's transformation

Want to make the periodic continuation continuous.

If $f(0) \neq f(1)$, define $\tilde{f}$ by $\tilde{f}(1-u) = \tilde{f}(u) = f(2u)$ for $0 \leq u \leq 1/2$.
This $\tilde{f}$ has the same integral as f and $\tilde{f}(0) = \tilde{f}(1)$.

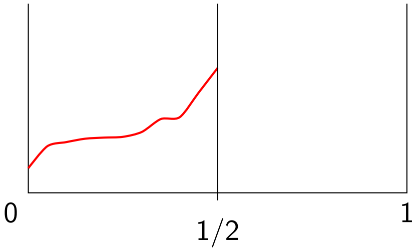


.

Baker's transformation

Want to make the periodic continuation continuous.

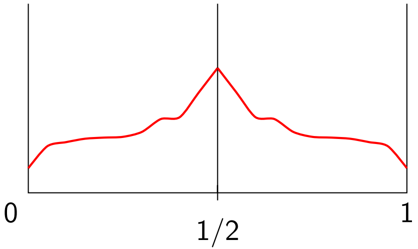
If $f(0) \neq f(1)$, define $\tilde{f}$ by $\tilde{f}(1-u) = \tilde{f}(u) = f(2u)$ for $0 \leq u \leq 1/2$.
This $\tilde{f}$ has the same integral as f and $\tilde{f}(0) = \tilde{f}(1)$.



Baker's transformation

Want to make the periodic continuation continuous.

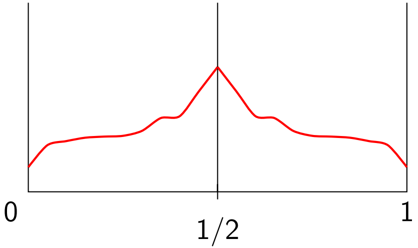
If $f(0) \neq f(1)$, define $\tilde{f}$ by $\tilde{f}(1-u) = \tilde{f}(u) = f(2u)$ for $0 \leq u \leq 1/2$.
This $\tilde{f}$ has the same integral as f and $\tilde{f}(0) = \tilde{f}(1)$.



Baker's transformation

Want to make the periodic continuation continuous.

If $f(0) \neq f(1)$, define $\tilde{f}$ by $\tilde{f}(1-u) = \tilde{f}(u) = f(2u)$ for $0 \leq u \leq 1/2$.
This $\tilde{f}$ has the same integral as f and $\tilde{f}(0) = \tilde{f}(1)$.

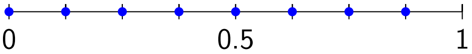


For smooth f, can reduce the variance to $O(n^{-4+\delta})$ (Hickernell 2002).
The resulting $\tilde{f}$ also symmetric with respect to $u = 1/2$.

In practice, we transform the points $\mathbf{U}_i$ instead of f.

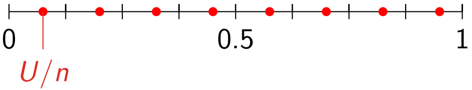
One-dimensional case

Random shift followed by baker's transformation.
Along each coordinate, stretch everything by a factor of 2 and fold.
Same as replacing U_j by $\min[2U_j, 2(1-U_j)]$.



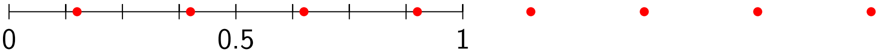
One-dimensional case

Random shift followed by baker's transformation.
Along each coordinate, stretch everything by a factor of 2 and fold.
Same as replacing U_j by $\min[2U_j, 2(1 - U_j)]$.



One-dimensional case

Random shift followed by baker's transformation.
Along each coordinate, stretch everything by a factor of 2 and fold.
Same as replacing U_j by $\min[2U_j, 2(1-U_j)]$.



One-dimensional case

Random shift followed by baker's transformation.
Along each coordinate, stretch everything by a factor of 2 and fold.
Same as replacing U_j by $\min[2U_j, 2(1-U_j)]$.
Gives locally antithetic points in intervals of size $2/n$.

Searching for a lattice that minimizes

$$\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} |\hat{f}(\mathbf{h})|^2$$

is impractical, because:

- the Fourier coefficients are usually unknown,
- there are infinitely many,
- must do it for each f.

ANOVA decomposition

The Fourier expansion has too many terms to handle. As a cruder expansion, we can write $f(\mathbf{u}) = f(u_1, \dots, u_s)$ as:

$$f(\mathbf{u}) = \sum_{\mathfrak{u} \subseteq \{1,\dots,s\}} f_{\mathfrak{u}}(\mathbf{u}) = \mu + \sum_{i=1}^{s} f_{\{i\}}(u_i) + \sum_{i,j=1}^{s} f_{\{i,j\}}(u_i, u_j) + \cdots$$

where

$$f_{\mathfrak{u}}(\mathbf{u}) = \int_{[0,1)^{|\bar{\mathfrak{u}}|}} f(\mathbf{u})\, \mathrm{d}\mathbf{u}_{\bar{\mathfrak{u}}} - \sum_{\mathfrak{v} \subset \mathfrak{u}} f_{\mathfrak{v}}(\mathbf{u}_{\mathfrak{v}}),$$

and the Monte Carlo variance decomposes as

$$\sigma^2 = \sum_{\mathfrak{u} \subseteq \{1,\dots,s\}} \sigma_{\mathfrak{u}}^2,$$

where the $\sigma_{\mathfrak{u}}^2 = \mathrm{Var}[f_{\mathfrak{u}}(\mathbf{U})]$ can be estimated by MC or RQMC.

Heuristic intuition: Make sure the projections of P_n are very uniform for the important subsets $\mathfrak{u}$ (i.e., with large $\sigma_{\mathfrak{u}}^2$).

Regrouping by projections: projection weights

$$\mathrm{Var}[\hat\mu_{n,\mathrm{rqmc}}] = \sum_{\mathfrak{u}\subseteq\{1,\dots,s\}} \mathrm{Var}[\hat\mu_{n,\mathrm{rqmc}}(f_{\mathfrak{u}})].$$

Denote $\mathfrak{u}(\mathbf{h}) = \mathfrak{u}(h_1,\dots,h_s)$ the set of indices j for which $h_j \neq 0$.
We will search for a lattice that minimizes the weighted $\mathcal{P}_{2\alpha}$:

$$\begin{aligned}
\mathcal{P}_{\gamma,2\alpha}(P_n^0) &= \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} \gamma_{\mathfrak{u}(\mathbf{h})}(\max(1,h_1),\dots,\max(1,h_s))^{-2\alpha} \\
&= \sum_{\emptyset\neq\mathfrak{u}\subseteq\{1,\dots,s\}} \frac{1}{n}\sum_{i=0}^{n-1}\gamma_{\mathfrak{u}}\left[\frac{-(-4\pi^2)^{\alpha}}{(2\alpha)!}\right]^{|\mathfrak{u}|}\prod_{j\in\mathfrak{u}} B_{2\alpha}(u_{i,j}),
\end{aligned}$$

where the projection-dependent weights $\gamma_{\mathfrak{u}}$ are positive real numbers, α is a positive integer, $\mathbf{u}_i = (u_{i,1},\dots,u_{i,s}) = i\mathbf{v}_1 \bmod 1$, $|\mathfrak{u}|$ is the cardinality of $\mathfrak{u}$, and $B_{2\alpha}$ is the Bernoulli polynomial of order 2α.

How to select the weights?

ANOVA variance components for worst-case $f_{\mathfrak{u}}$, whose square Fourier coefficients are $|\hat{f}(\mathbf{h})|^2 = \max(1, h_1), \ldots, \max(1, h_s))^{-2\alpha}$:

$$\sigma_{\mathfrak{u}}^2 = \gamma_{\mathfrak{u}} \left[|B_{2\alpha}(0)| \frac{(4\pi^2)^\alpha}{(2\alpha)!} \right]^{|\mathfrak{u}|} \stackrel{\text{def}}{=} \gamma_{\mathfrak{u}} (\kappa(\alpha))^{-|\mathfrak{u}|}$$

where $\kappa(\alpha)$ depends on α. We have

$$\kappa(1) = \frac{3}{\pi^2} \approx 0.30396, \quad \kappa(2) = \frac{45}{\pi^4} \approx 0.46197, \quad \kappa(3) \approx 0.49148,$$

and $\kappa(\alpha) \to 0.5$ when $\alpha \to \infty$.

Idea: estimate the variance components $\sigma_{\mathfrak{u}}^2$ and take the weights

$$\gamma_{\mathfrak{u}} = \sigma_{\mathfrak{u}}^2 (\kappa(\alpha))^{|\mathfrak{u}|}.$$

Simplified choices of weights

Order-dependent weights:

$\gamma_{\mathfrak{u}}$ depends only on $|\mathfrak{u}|$.

Special case: $\gamma_{\mathfrak{u}} = 1$ for $|\mathfrak{u}| \leq d$ and $\gamma_{\mathfrak{u}} = 0$ otherwise.

Simplified choices of weights

Order-dependent weights:

$\gamma_{\mathfrak{u}}$ depends only on $|\mathfrak{u}|$.

Special case: $\gamma_{\mathfrak{u}} = 1$ for $|\mathfrak{u}| \leq d$ and $\gamma_{\mathfrak{u}} = 0$ otherwise.

Product weights:

$\gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_j$ for some constants $\gamma_j \geq 0$.

Geometric weights:

Take $\gamma_j = a\beta^j$ for $a, \beta > 0$.

Searching for lattice parameters

Korobov lattices.

Search for $\mathbf{z} = (1, a, a^2, \ldots, ...)$ over all admissible integers a.

Searching for lattice parameters

Korobov lattices.

Search for $\mathbf{z} = (1, a, a^2, \ldots, ...)$ over all admissible integers a.

Component by component (CBC) construction.

Let $z_1 = 1$;

For $j = 2, \ldots, s$, find $z_j \in \{1, \ldots, n-1\}$, $\gcd(z_j, n) = 1$, such that $(z_1, \ldots, z_{j-1}, z_j)$ minimizes the selected discrepancy for the first j dimensions.

Searching for lattice parameters

Korobov lattices.

Search for $\mathbf{z} = (1, a, a^2, \ldots, ...)$ over all admissible integers a.

Component by component (CBC) construction.

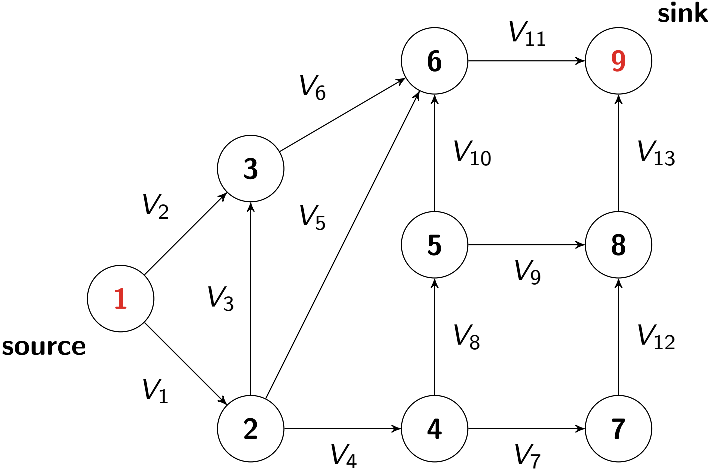
> Let $z_1 = 1$;
> For $j = 2, \ldots, s$, find $z_j \in \{1, \ldots, n-1\}$, $\gcd(z_j, n) = 1$, such that $(z_1, \ldots, z_{j-1}, z_j)$ minimizes the selected discrepancy for the first j dimensions.

Partial randomized CBC construction.

> Let $z_1 = 1$;
> For $j = 2, \ldots, s$, try r random $z_j \in \{1, \ldots, n-1\}$, $\gcd(z_j, n) = 1$, and retain the one for which $(z_1, \ldots, z_{j-1}, z_j)$ minimizes the selected discrepancy for the first j dimensions.

Example: stochastic activity network

Each arc j has random length $V_j = F_j^{-1}(U_j)$.
Let $T = f(U_1, \ldots, U_{13})$ = length of longest path from node 1 to node 9.
Want to estimate $q(x) = \mathbb{P}[T > x]$ for a given constant x.



To estimate $q(x)$ by **MC**, we generate n independent realizations of T, say $T_1, \ldots, T_n$, and take $(1/n) \sum_{i=1}^{n} \mathbb{I}[T_i > x]$.

For **RQMC**, we replace the n realizations of $(U_1, \ldots, U_{13})$ by the n points of a randomly-shifted lattice.

To estimate $q(x)$ by **MC**, we generate n independent realizations of T, say $T_1, \ldots, T_n$, and take $(1/n)\sum_{i=1}^{n} \mathbb{I}[T_i > x]$.

For **RQMC**, we replace the n realizations of $(U_1, \ldots, U_{13})$ by the n points of a randomly-shifted lattice.

Illustration: $V_j \sim \text{Normal}(\mu_j, \sigma_j^2)$ for $j = 1, 2, 4, 11, 12$, and $V_j \sim \text{Exponential}(1/\mu_j)$ otherwise.

The μ_j: 13.0, 5.5, 7.0, 5.2, 16.5, 14.7, 10.3, 6.0, 4.0, 20.0, 3.2, 3.2, 16.5.

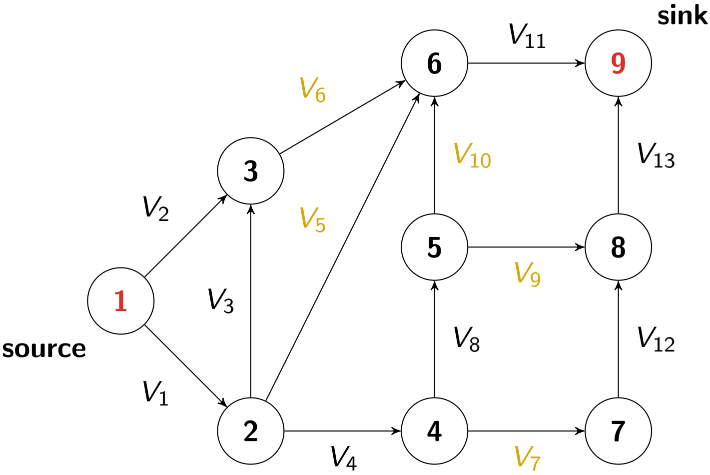
To estimate $q(x)$ by **MC**, we generate n independent realizations of T, say $T_1, \ldots, T_n$, and take $(1/n)\sum_{i=1}^{n} \mathbb{I}[T_i > x]$.

For **RQMC**, we replace the n realizations of $(U_1, \ldots, U_{13})$ by the n points of a randomly-shifted lattice.

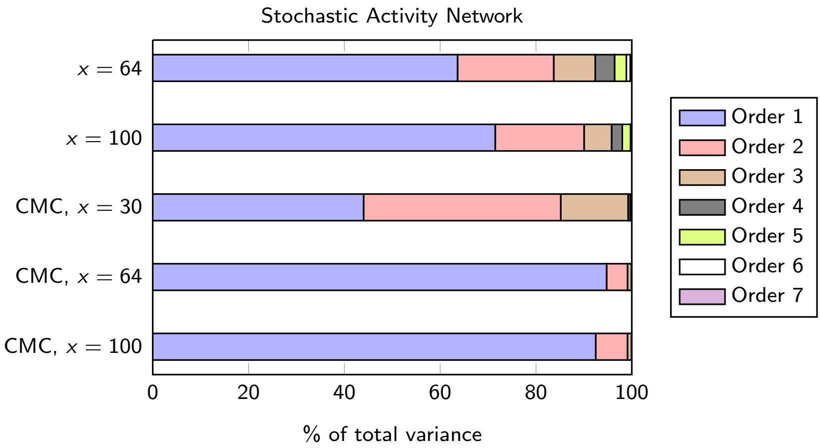
Illustration: $V_j \sim \text{Normal}(\mu_j, \sigma_j^2)$ for $j = 1, 2, 4, 11, 12$, and $V_j \sim \text{Exponential}(1/\mu_j)$ otherwise.

The μ_j: 13.0, 5.5, 7.0, 5.2, 16.5, 14.7, 10.3, 6.0, 4.0, 20.0, 3.2, 3.2, 16.5.

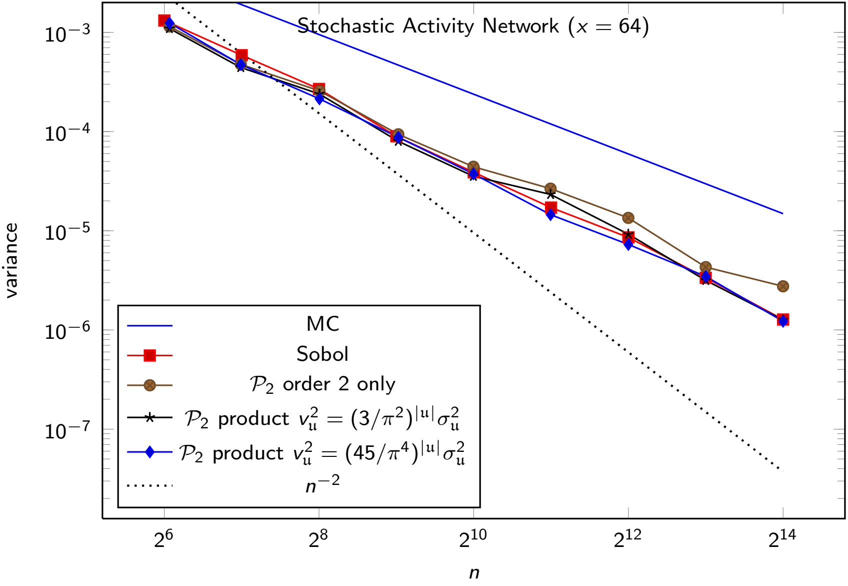
CMC estimator. Generate the V_j's only for the 8 arcs that do not belong to the cut $\mathcal{L} = \{5, 6, 7, 9, 10\}$, and replace $\mathbb{I}[T > x]$ by its conditional expectation given those V_j's, $\mathbb{P}[T > x \mid \{V_j,\, j \notin \mathcal{L}\}]$.
This makes the integrand continuous in the U_j's.


sink
source
1
2
3
4
5
6
7
8
9
V_1
V_2
V_3
V_4
V_5
V_6
V_7
V_8
V_9
V_{10}
V_{11}
V_{12}
V_{13}

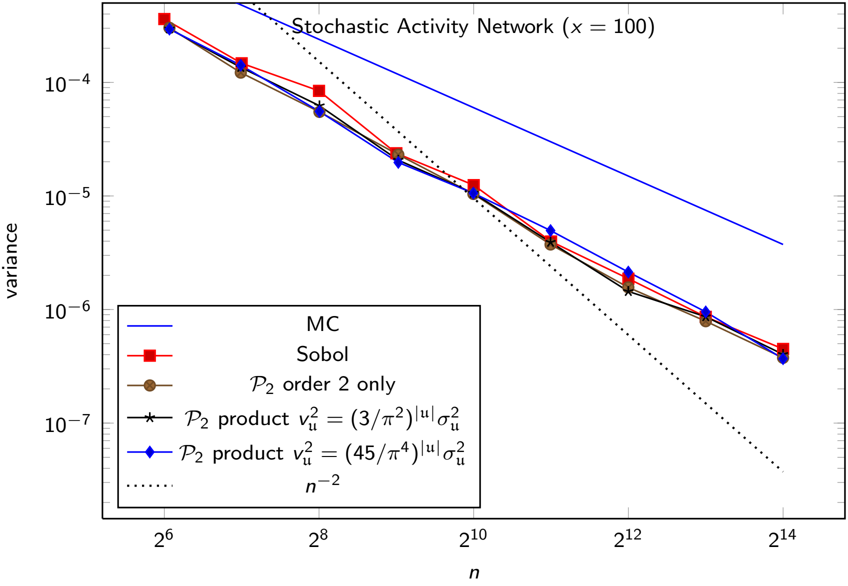
ANOVA Variances for the Stochastic Activity Network



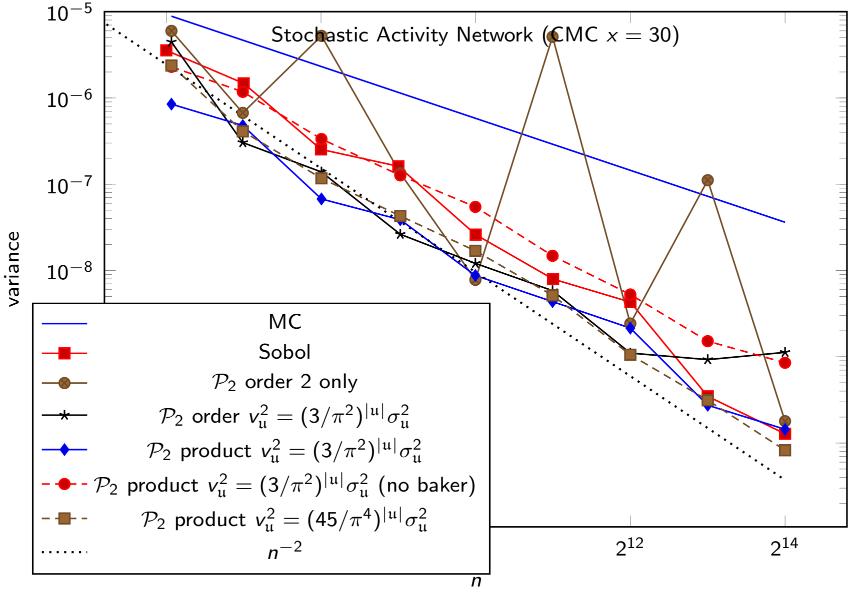
Lattices of Rank 1 with CBC



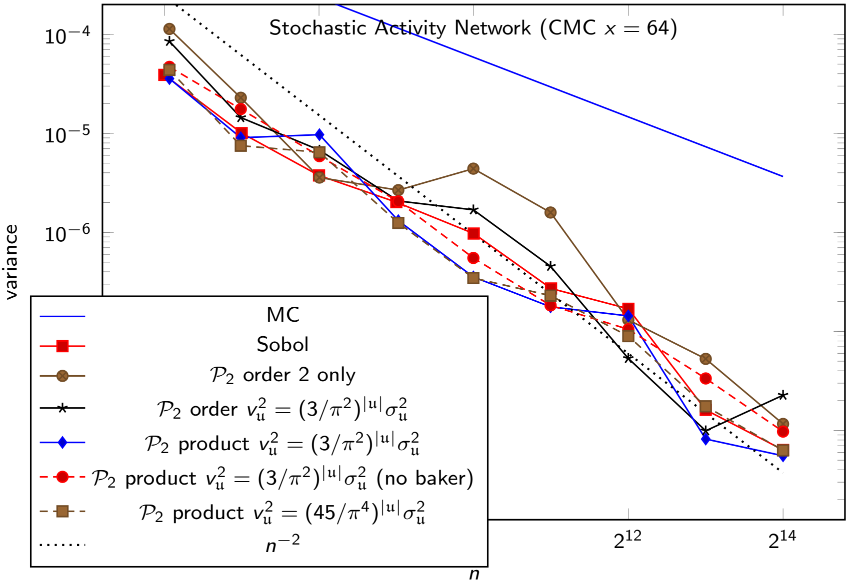
Lattices of Rank 1 with CBC



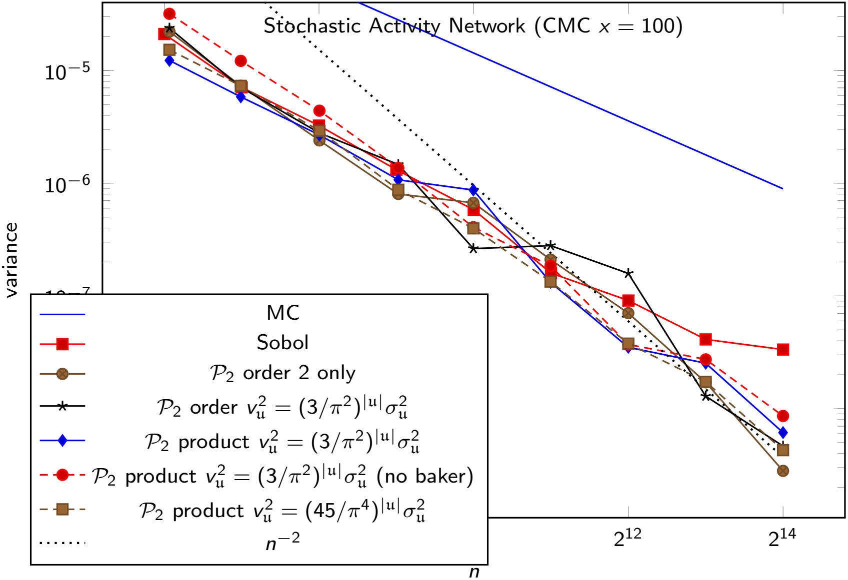
Lattices of Rank 1 with CBC



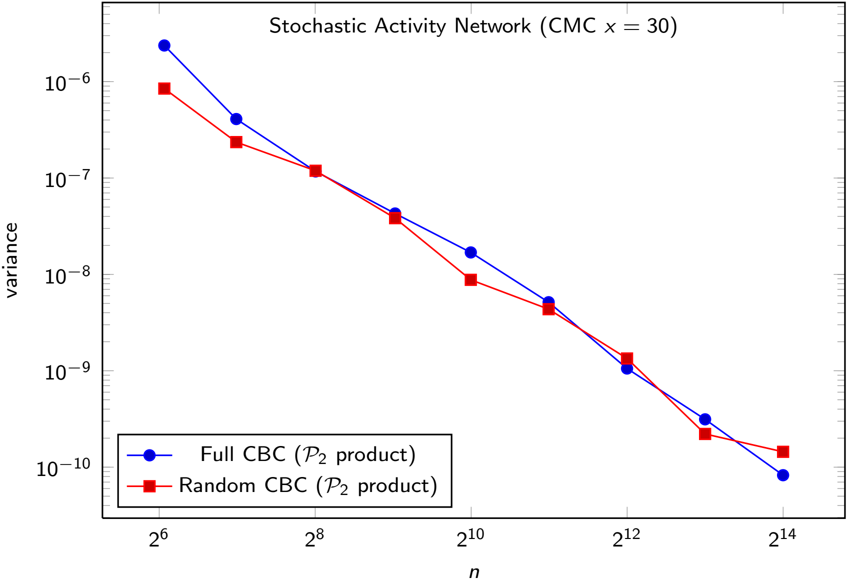
Lattices of Rank 1 with CBC



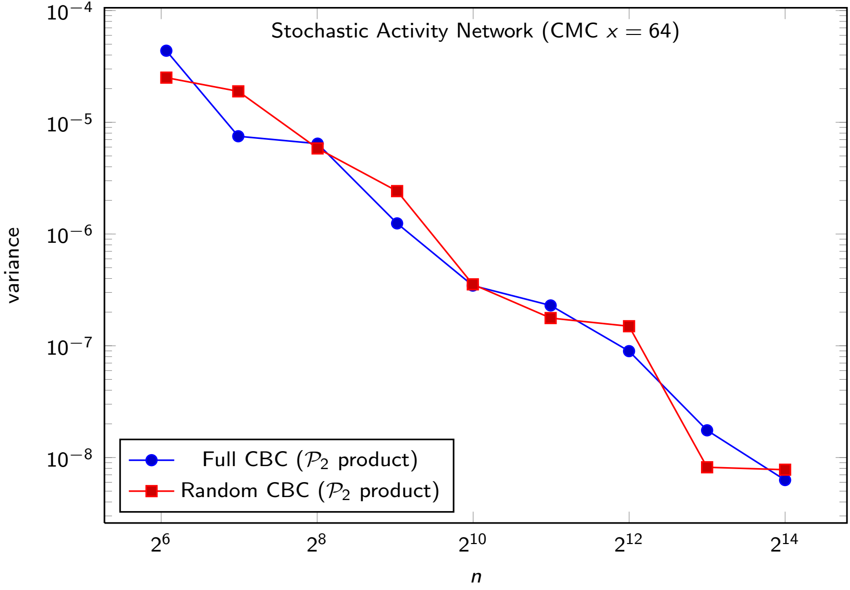
Lattices of Rank 1 with CBC



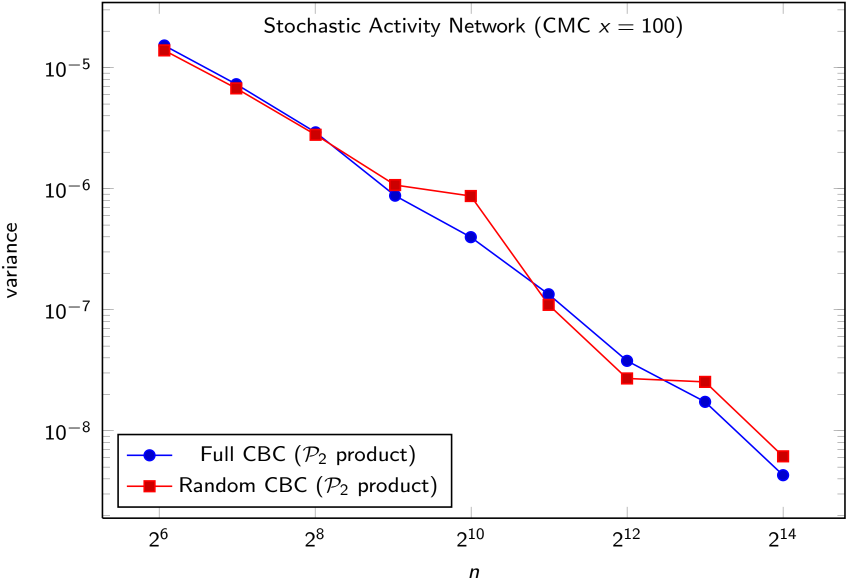
Random vs. Full CBC



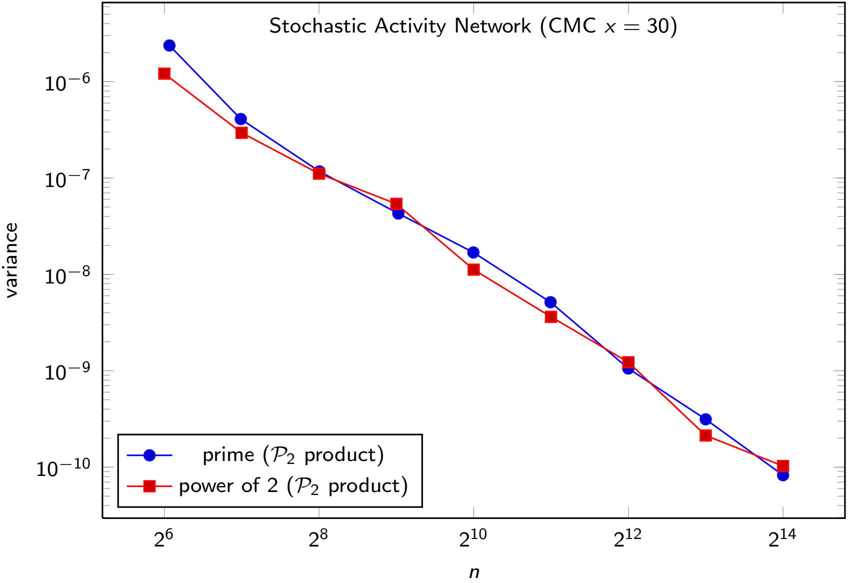
Random vs. Full CBC



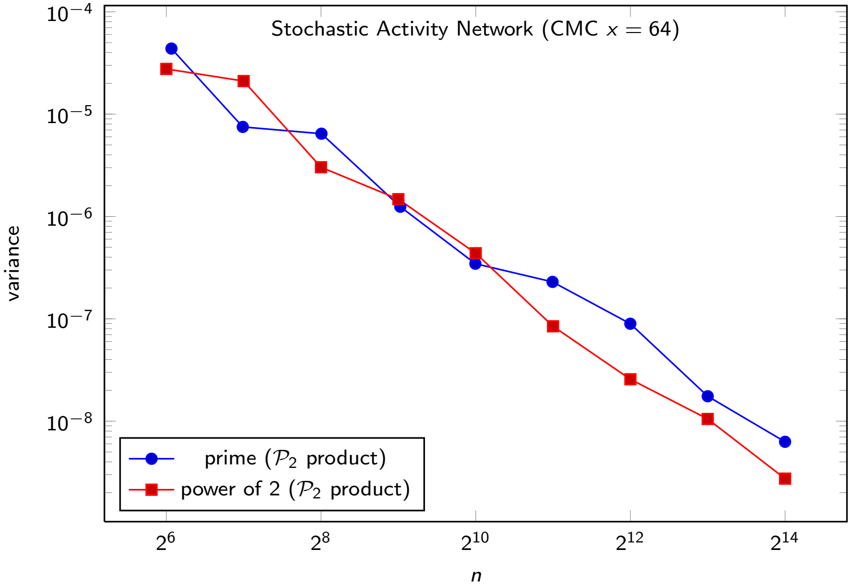
Random vs. Full CBC



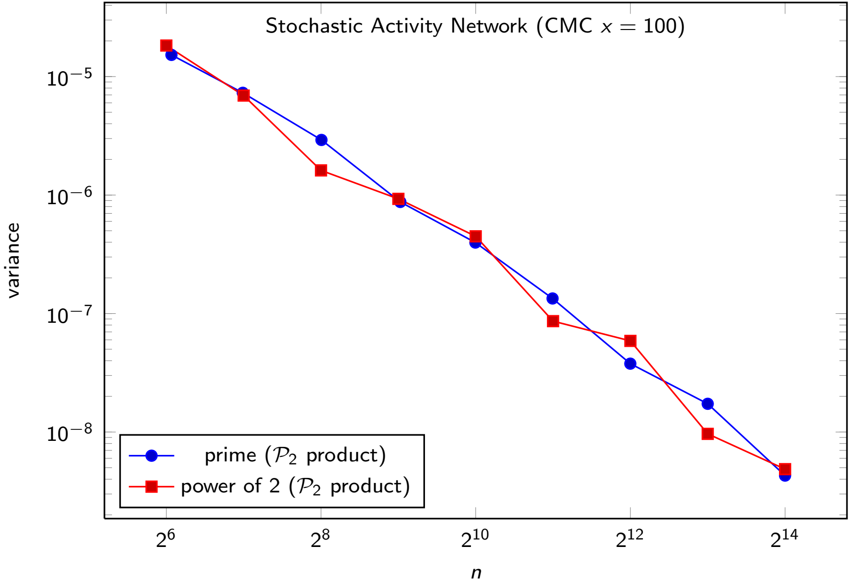
Prime vs. Power-of-2 Number of Points



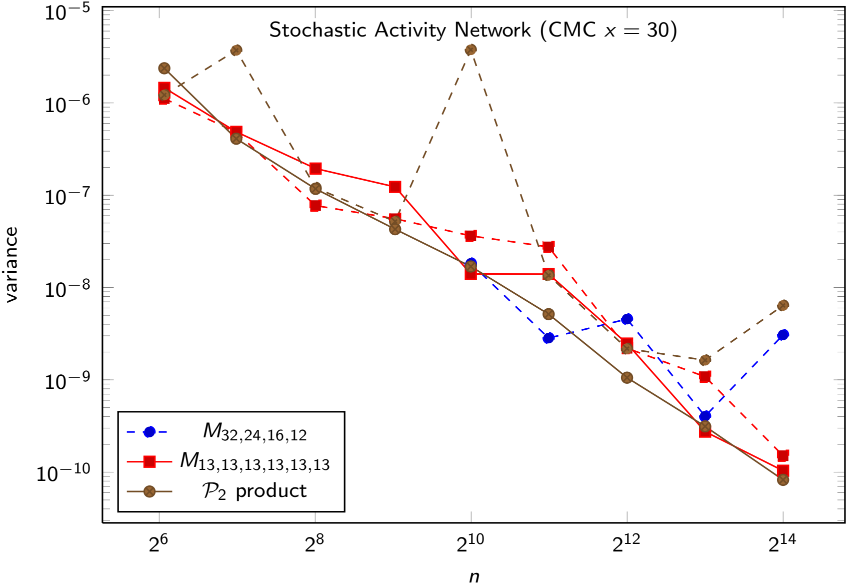
Prime vs. Power-of-2 Number of Points



Prime vs. Power-of-2 Number of Points

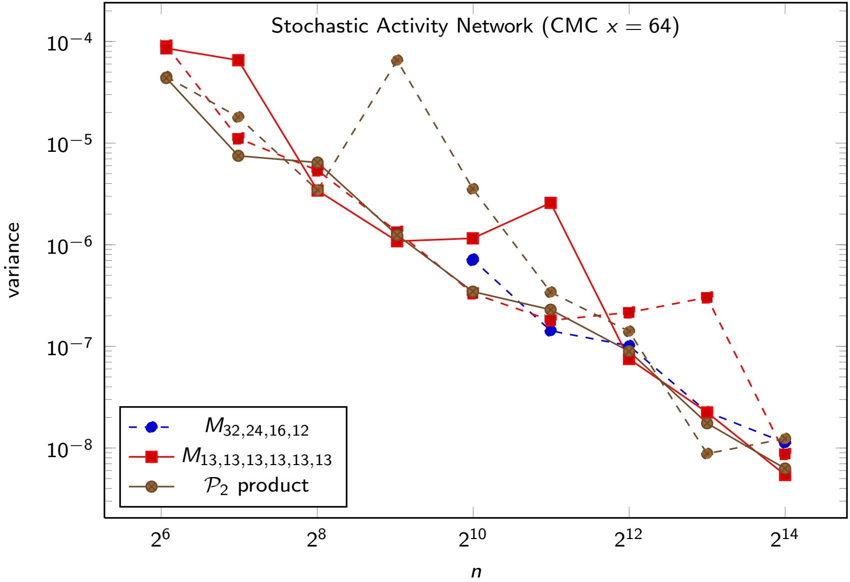


Korobov vs. CBC



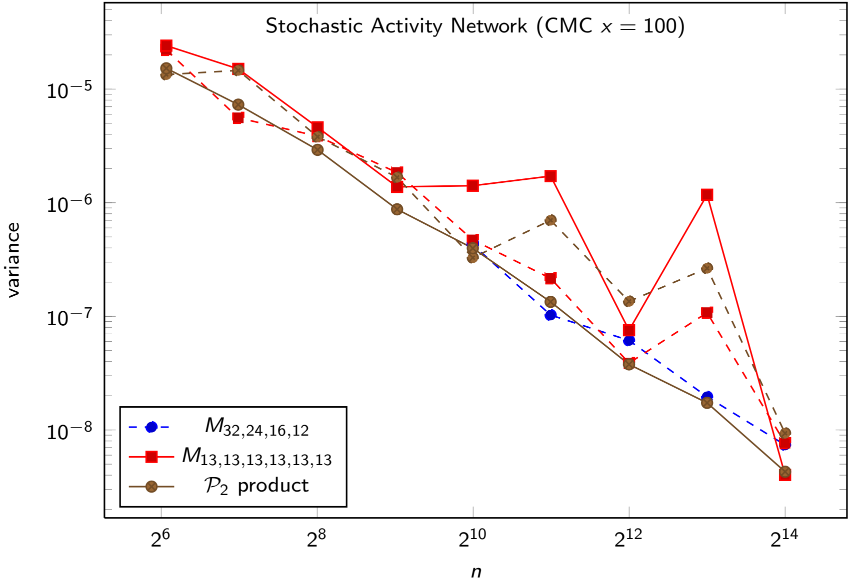
Solid: CBC. Dashed: Korobov.

Korobov vs. CBC



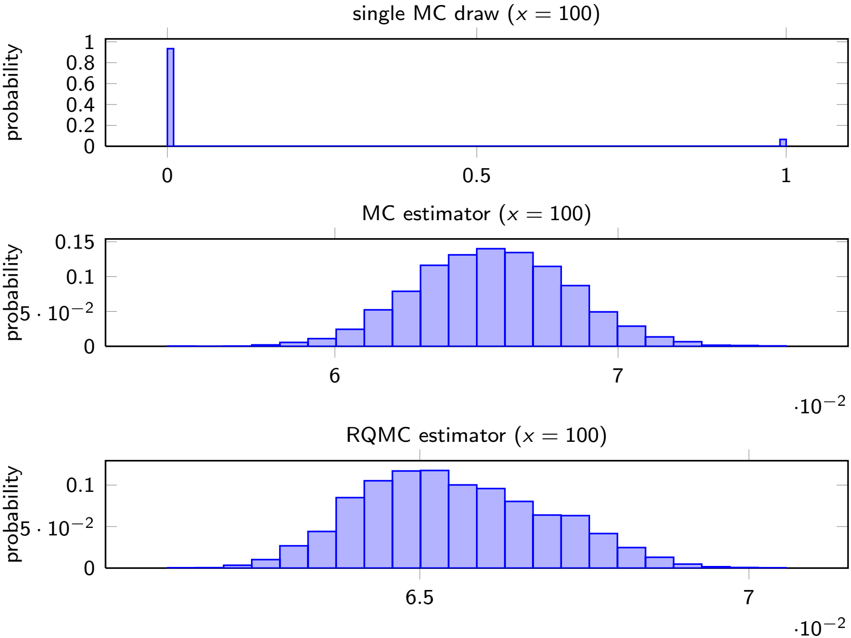
Solid: CBC. Dashed: Korobov.

Korobov vs. CBC

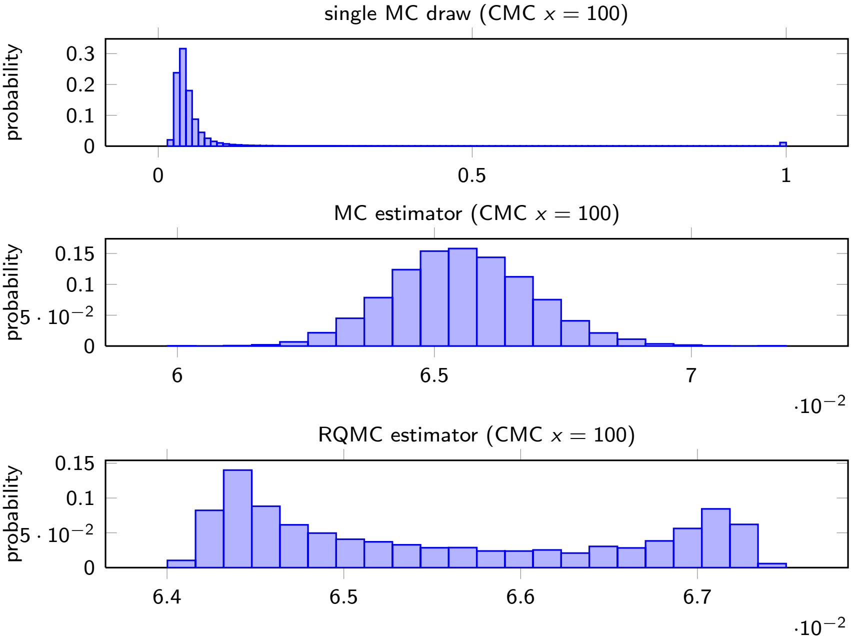


Solid: CBC. Dashed: Korobov.

Histograms, for $n = 8191$, $m = 10^4$ replications



Histograms


single MC draw (CMC $x = 100$)
probability
0.3
0.2
0.1
0
0
0.5
1
MC estimator (CMC $x = 100$)
probability
0.15
0.1
$5 \cdot 10^{-2}$
0
6
6.5
7
$\cdot 10^{-2}$
RQMC estimator (CMC $x = 100$)
probability
0.15
0.1
$5 \cdot 10^{-2}$
0
6.4
6.5
6.6
6.7
$\cdot 10^{-2}$

Function of a Multinormal vector

Let $\mu = E[f(\mathbf{U})] = E[g(\mathbf{Y})]$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \mathbf{\Sigma})$.

Function of a Multinormal vector

Let $\mu = E[f(\mathbf{U})] = E[g(\mathbf{Y})]$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

For example, if the payoff of a financial derivative is a function of the values taken by a c-dimensional geometric Brownian motions (GMB) at d observations times $0 < t_1 < \cdots < t_d = T$, then we have $s = cd$.

Function of a Multinormal vector

Let $\mu = E[f(\mathbf{U})] = E[g(\mathbf{Y})]$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

For example, if the payoff of a financial derivative is a function of the values taken by a c-dimensional geometric Brownian motions (GMB) at d observations times $0 < t_1 < \cdots < t_d = T$, then we have $s = cd$.

To generate $\mathbf{Y}$: Decompose $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{t}}$, generate
$\mathbf{Z} = (Z_1, \ldots, Z_s) = (\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_s)) \sim N(\mathbf{0}, \mathbf{I})$
and return $\mathbf{Y} = \mathbf{A}\mathbf{Z}$.

Function of a Multinormal vector

Let $\mu = E[f(\mathbf{U})] = E[g(\mathbf{Y})]$ where $\mathbf{Y} = (Y_1, \dots, Y_s) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

For example, if the payoff of a financial derivative is a function of the values taken by a c-dimensional geometric Brownian motions (GMB) at d observations times $0 < t_1 < \cdots < t_d = T$, then we have $s = cd$.

To generate $\mathbf{Y}$: Decompose $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{t}}$, generate
$\mathbf{Z} = (Z_1, \dots, Z_s) = (\Phi^{-1}(U_1), \dots, \Phi^{-1}(U_s)) \sim N(\mathbf{0}, \mathbf{I})$
and return $\mathbf{Y} = \mathbf{A}\mathbf{Z}$.
Choice of $\mathbf{A}$?

Function of a Multinormal vector

Let $\mu = E[f(\mathbf{U})] = E[g(\mathbf{Y})]$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \mathbf{\Sigma})$.

For example, if the payoff of a financial derivative is a function of the values taken by a c-dimensional geometric Brownian motions (GMB) at d observations times $0 < t_1 < \cdots < t_d = T$, then we have $s = cd$.

To generate $\mathbf{Y}$: Decompose $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{t}}$, generate
$\mathbf{Z} = (Z_1, \ldots, Z_s) = (\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_s)) \sim N(\mathbf{0}, \mathbf{I})$
and return $\mathbf{Y} = \mathbf{A}\mathbf{Z}$.
Choice of $\mathbf{A}$?

Cholesky factorization: $\mathbf{A}$ is lower triangular.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\mathbf{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this $\mathbf{A}$, Z_1 accounts for the maximum amount of variance of $\mathbf{Y}$, then Z_2 for the maximum amount of variance conditional on Z_1, and so on.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this $\mathbf{A}$, Z_1 accounts for the maximum amount of variance of $\mathbf{Y}$, then Z_2 for the maximum amount of variance conditional on Z_1, and so on.

Function of a Brownian motion:
Payoff depends on c-dimensional Brownian motion $\{\mathbf{X}(t),\ t \geq 0\}$ observed at times $0 = t_0 < t_1 < \cdots < t_d$.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this $\mathbf{A}$, Z_1 accounts for the maximum amount of variance of $\mathbf{Y}$, then Z_2 for the maximum amount of variance conditional on Z_1, and so on.

Function of a Brownian motion:
Payoff depends on c-dimensional Brownian motion $\{\mathbf{X}(t),\ t \geq 0\}$ observed at times $0 = t_0 < t_1 < \cdots < t_d$.

Sequential (or random walk) method: generate $\mathbf{X}(t_1)$, then $\mathbf{X}(t_2) - \mathbf{X}(t_1)$, then $\mathbf{X}(t_3) - \mathbf{X}(t_2)$, etc.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this $\mathbf{A}$, Z_1 accounts for the maximum amount of variance of $\mathbf{Y}$, then Z_2 for the maximum amount of variance conditional on Z_1, and so on.

Function of a Brownian motion:
Payoff depends on c-dimensional Brownian motion $\{\mathbf{X}(t),\ t \geq 0\}$ observed at times $0 = t_0 < t_1 < \cdots < t_d$.

Sequential (or random walk) method: generate $\mathbf{X}(t_1)$, then $\mathbf{X}(t_2) - \mathbf{X}(t_1)$, then $\mathbf{X}(t_3) - \mathbf{X}(t_2)$, etc.

Brownian bridge (BB) sampling: Suppose $d = 2^m$.
Generate $\mathbf{X}(t_d)$, then $\mathbf{X}(t_{d/2})$ conditional on $(\mathbf{X}(0), \mathbf{X}(t_d))$,
then $\mathbf{X}(t_{d/4})$ conditional on $(\mathbf{X}(0), \mathbf{X}(t_{d/2}))$, and so on.

The first few $N(0,1)$ r.v.'s already sketch the path trajectory.

Principal component decomposition (PCA):
$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D} = \mathrm{diag}(\lambda_s, \ldots, \lambda_1)$ (eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order) and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this $\mathbf{A}$, Z_1 accounts for the maximum amount of variance of $\mathbf{Y}$, then Z_2 for the maximum amount of variance conditional on Z_1, and so on.

Function of a Brownian motion:
Payoff depends on c-dimensional Brownian motion $\{\mathbf{X}(t),\ t \geq 0\}$ observed at times $0 = t_0 < t_1 < \cdots < t_d$.

Sequential (or random walk) method: generate $\mathbf{X}(t_1)$, then $\mathbf{X}(t_2) - \mathbf{X}(t_1)$, then $\mathbf{X}(t_3) - \mathbf{X}(t_2)$, etc.

Brownian bridge (BB) sampling: Suppose $d = 2^m$.
Generate $\mathbf{X}(t_d)$, then $\mathbf{X}(t_{d/2})$ conditional on $(\mathbf{X}(0), \mathbf{X}(t_d))$,
then $\mathbf{X}(t_{d/4})$ conditional on $(\mathbf{X}(0), \mathbf{X}(t_{d/2}))$, and so on.

The first few $N(0,1)$ r.v.'s already sketch the path trajectory.

Each of these methods corresponds to some matrix $\mathbf{A}$.
Choice has large impact on the ANOVA decomposition of f.

Example: Pricing an Asian option

Single asset, s observation times $t_1, \ldots, t_s$. Want to estimate $\mathbb{E}[f(\mathbf{U})]$, where

$$f(\mathbf{U}) = e^{-rt_s} \max\left[0,\ \frac{1}{s}\sum_{j=1}^{s} S(t_j) - K\right]$$

and $\{S(t),\ t \geq 0\}$ is a geometric Brownian motion.
We have $f(\mathbf{U}) = g(\mathbf{Y})$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \mathbf{\Sigma})$.

Example: Pricing an Asian option

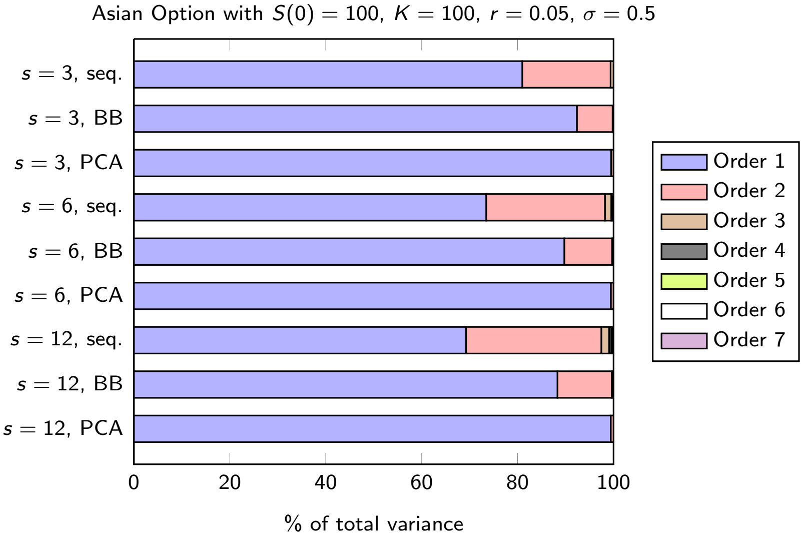
Single asset, s observation times $t_1, \ldots, t_s$. Want to estimate $\mathbb{E}[f(\mathbf{U})]$, where

$$f(\mathbf{U}) = e^{-rt_s} \max\left[0,\ \frac{1}{s}\sum_{j=1}^{s} S(t_j) - K\right]$$

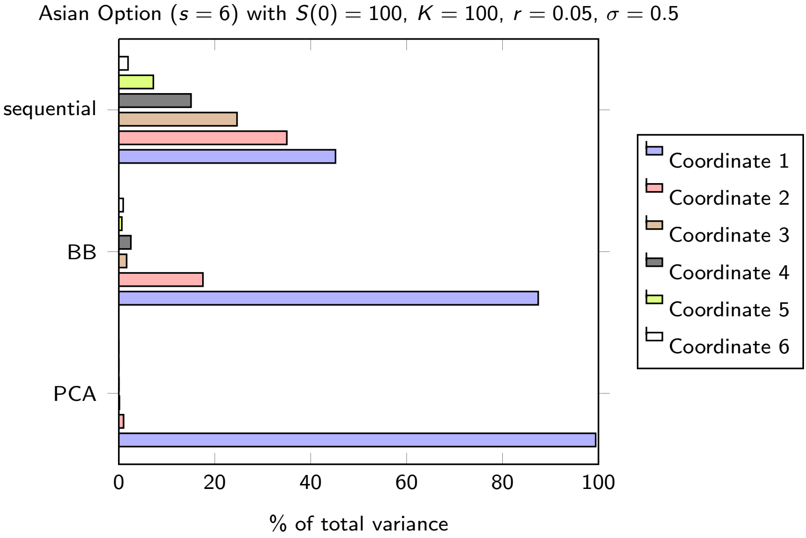
and $\{S(t),\ t \geq 0\}$ is a geometric Brownian motion.
We have $f(\mathbf{U}) = g(\mathbf{Y})$ where $\mathbf{Y} = (Y_1, \ldots, Y_s) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

Let $S(0) = 100$, $K = 100$, $r = 0.05$, $t_s = 1$, and $t_j = jT/s$ for $1 \leq j \leq s$.
We consider $\sigma = 0.2$, 0.5 and $s = 3, 6, 12$.

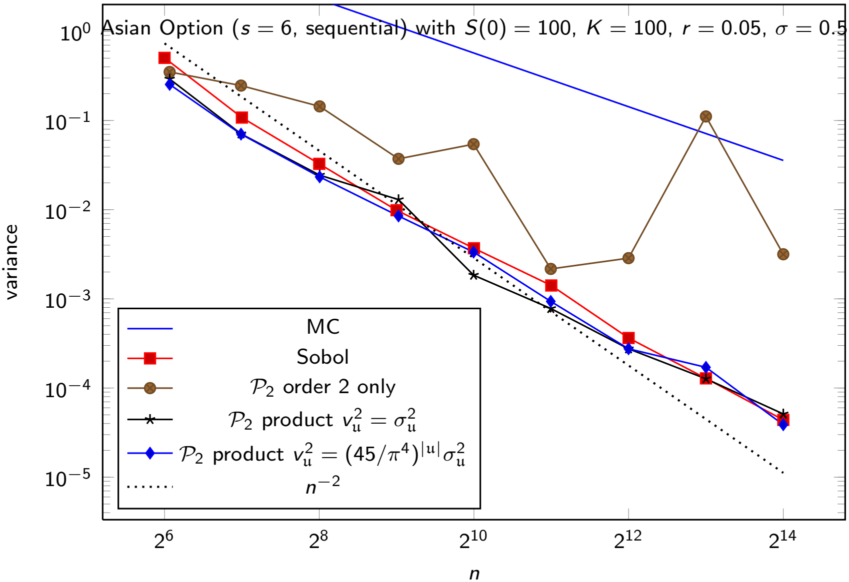
ANOVA Variances for the Asian Option



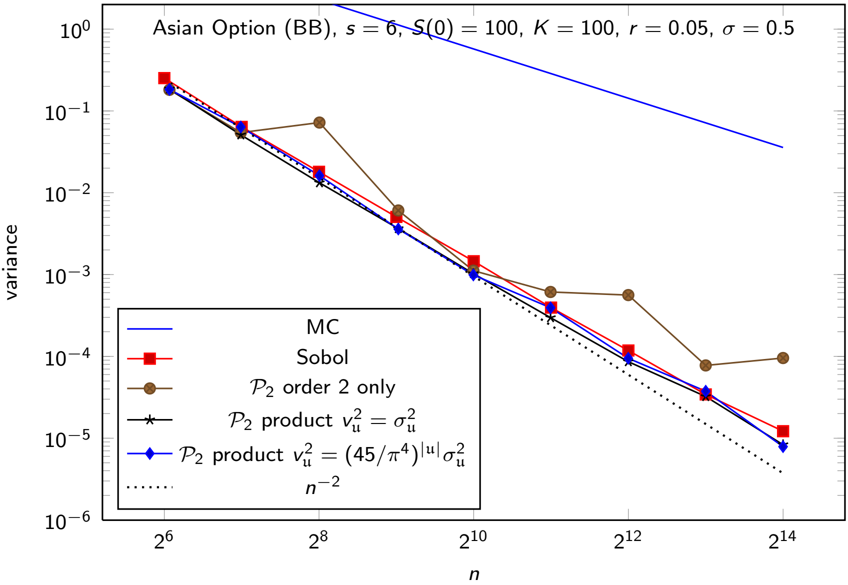
Total Variance per Coordinate for the Asian Option



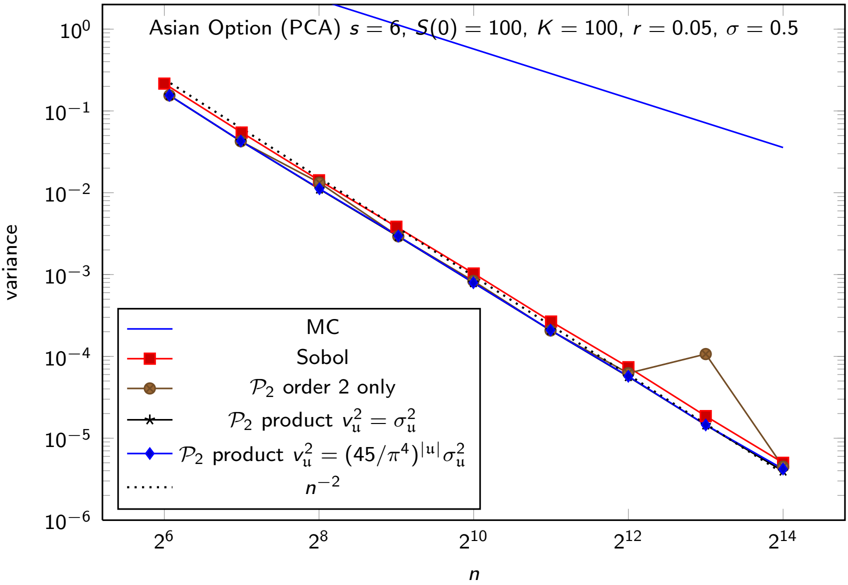
Lattices of Rank 1 with CBC



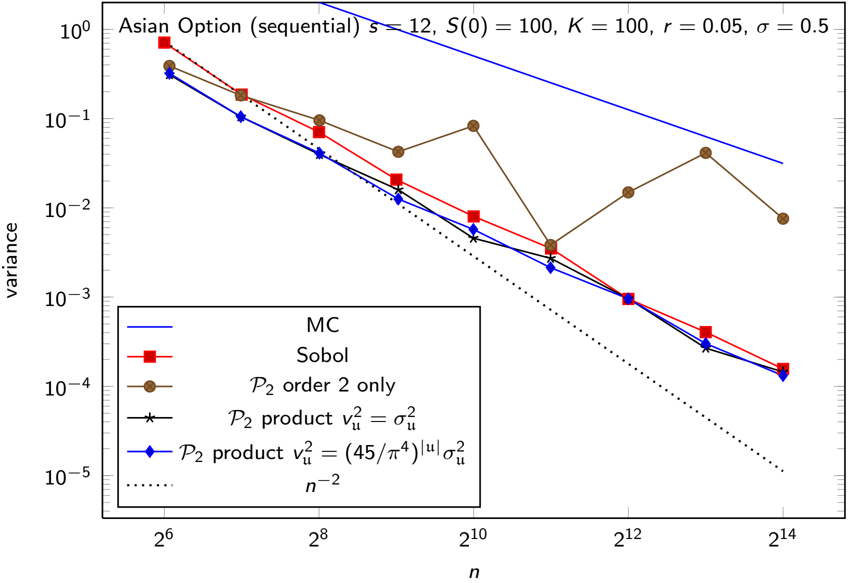
Lattices of Rank 1 with CBC



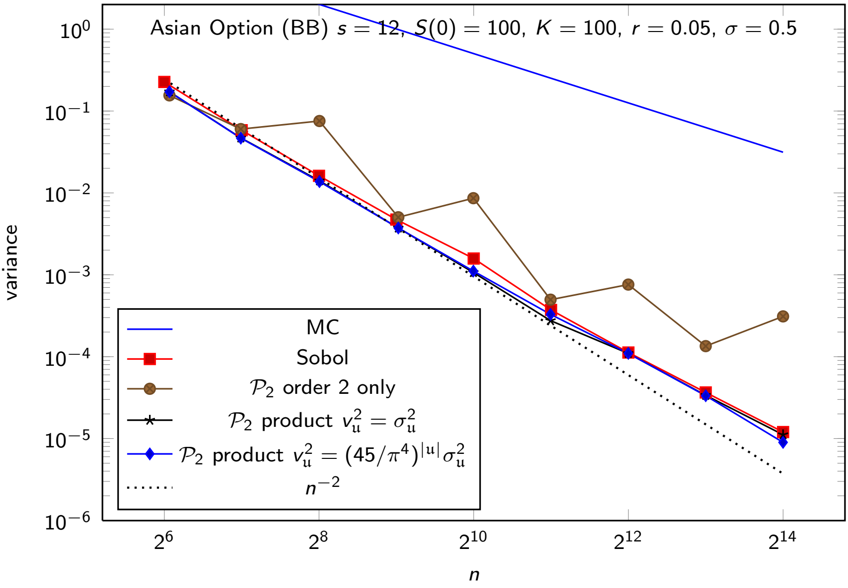
Lattices of Rank 1 with CBC



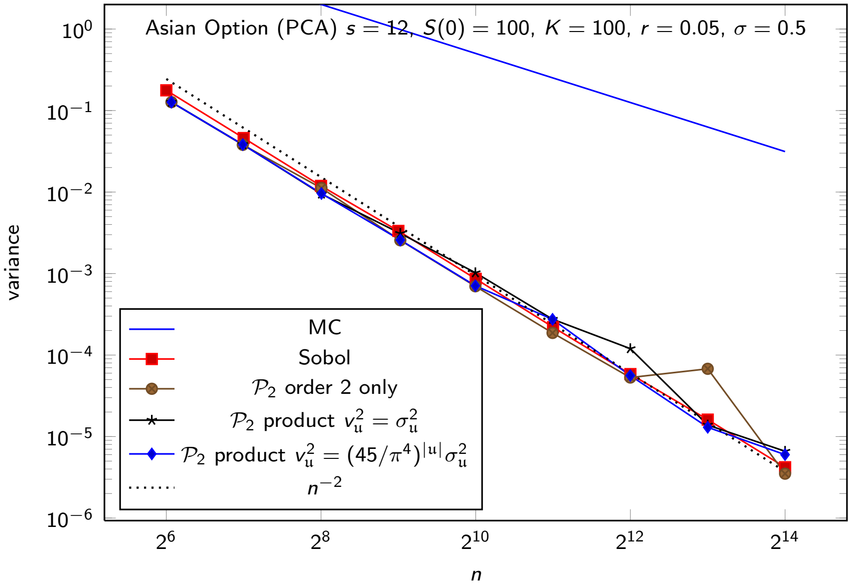
Lattices of Rank 1 with CBC



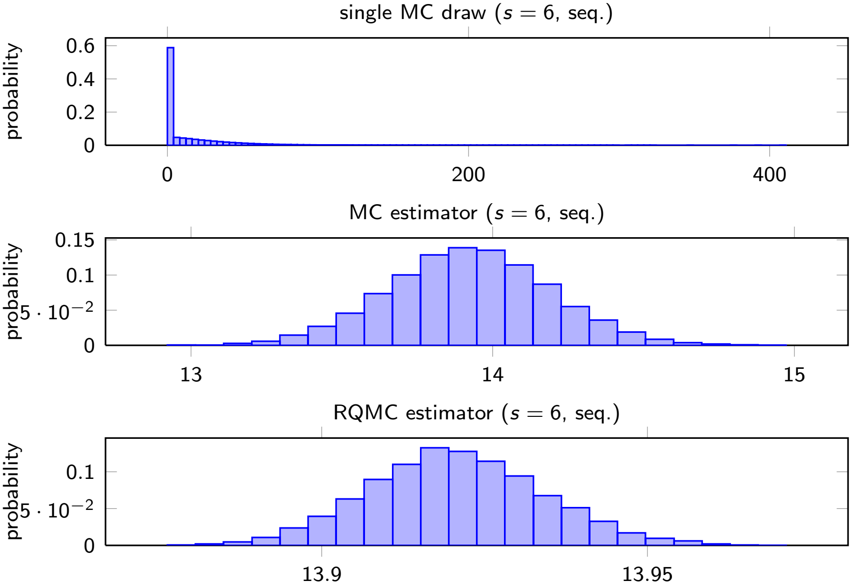
Lattices of Rank 1 with CBC



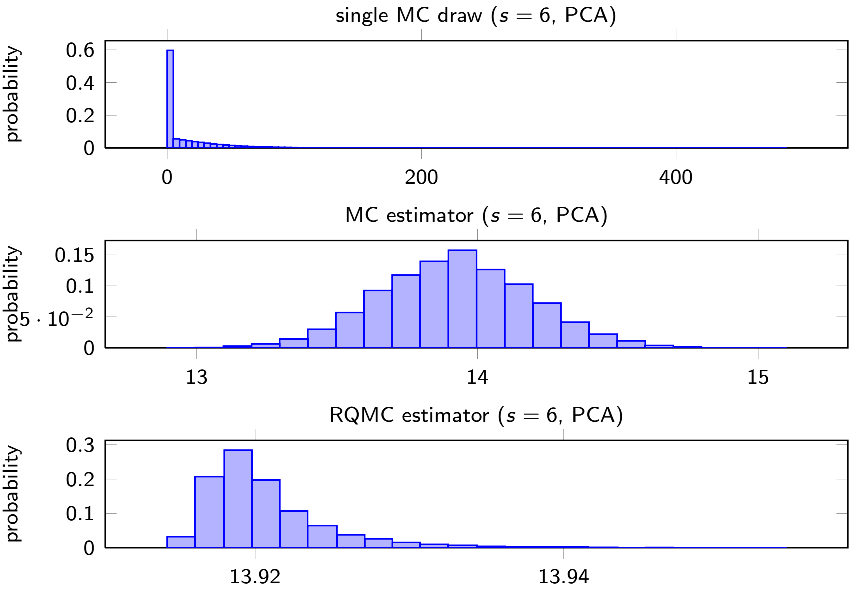
Lattices of Rank 1 with CBC



Histograms for the Asian Option, $s = 6$, sequential



Histograms for the Asian option, $s = 6$, PCA

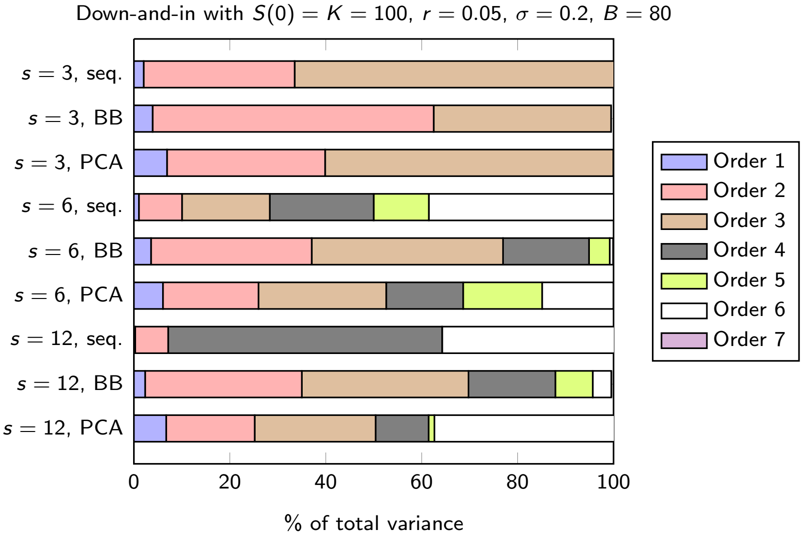


A down-and-in Asian option with barrier B

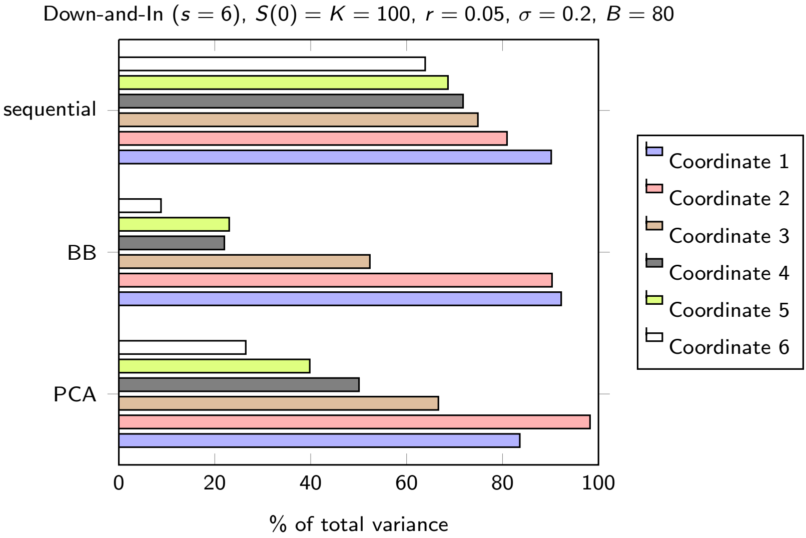
Same as for Asian option, except that payoff is zero unless

$$\min_{1 \le j \le s} S(t_j) \le 80.$$

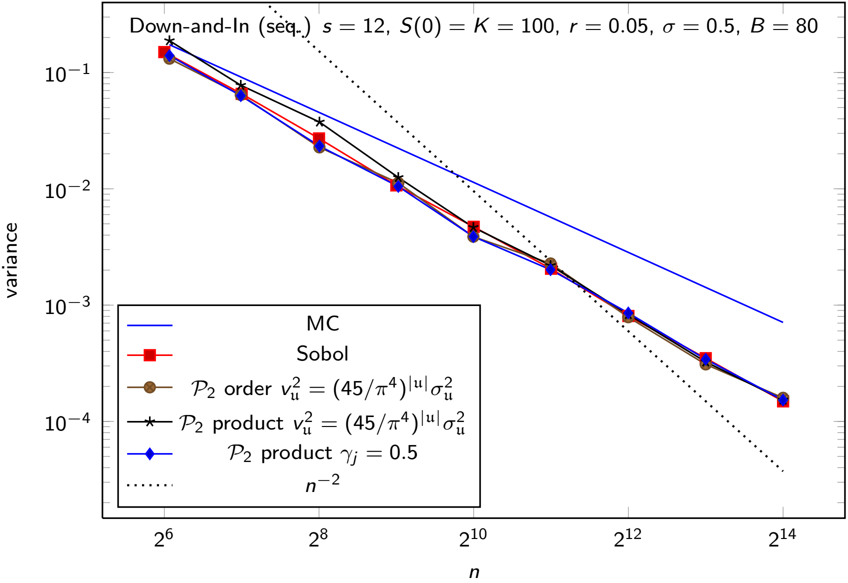
ANOVA Variances for the down-and-in Asian Option



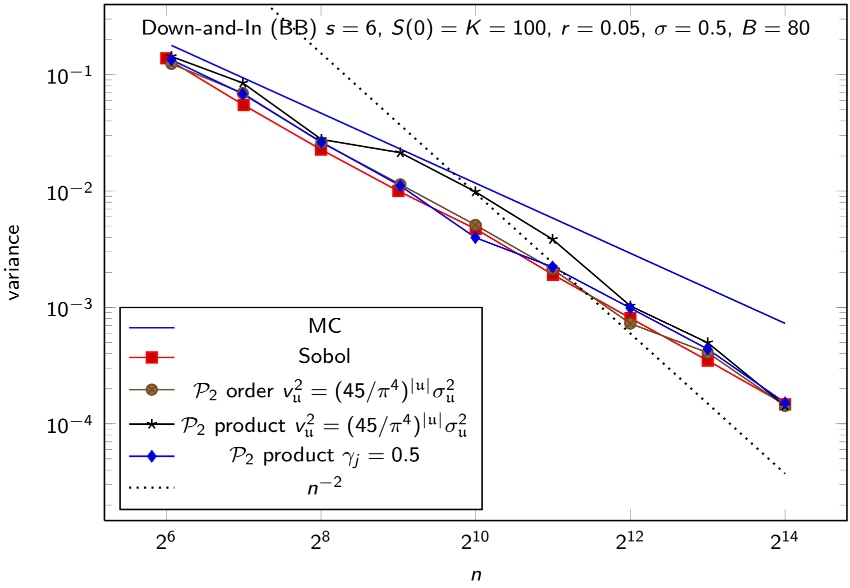
Total Variance per Coordinate for the down-and-in Asian Option



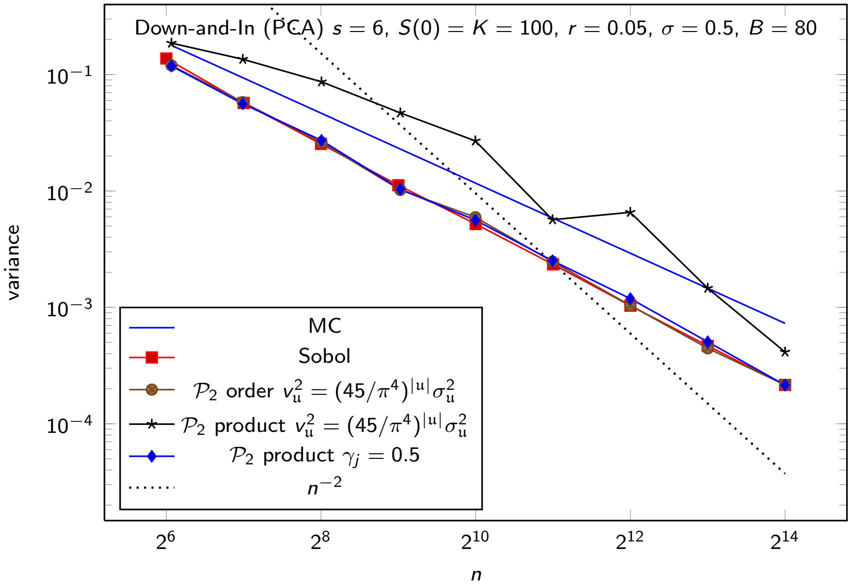
Lattices of Rank 1 with CBC



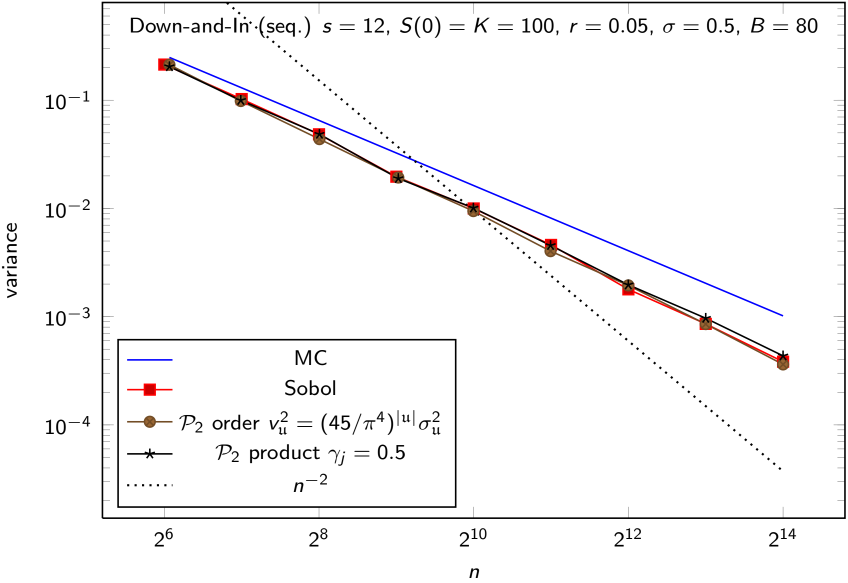
Lattices of Rank 1 with CBC



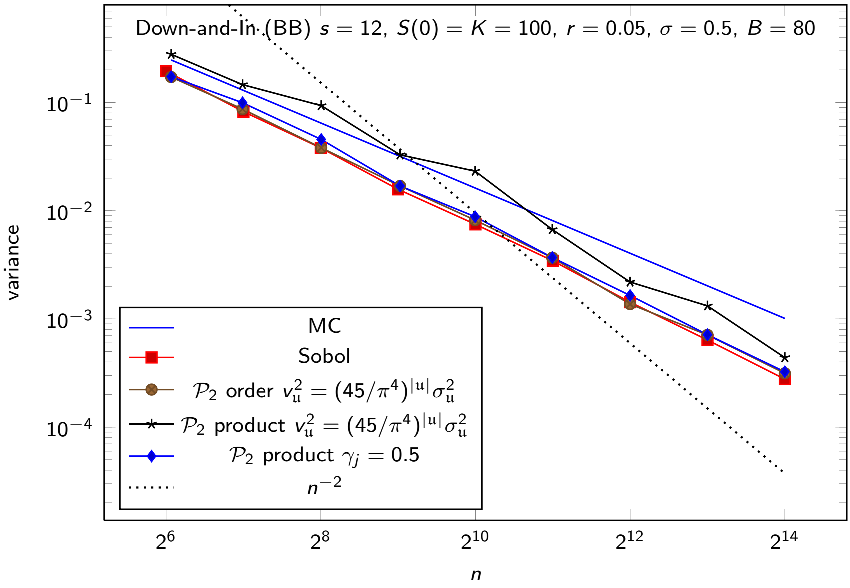
Lattices of Rank 1 with CBC



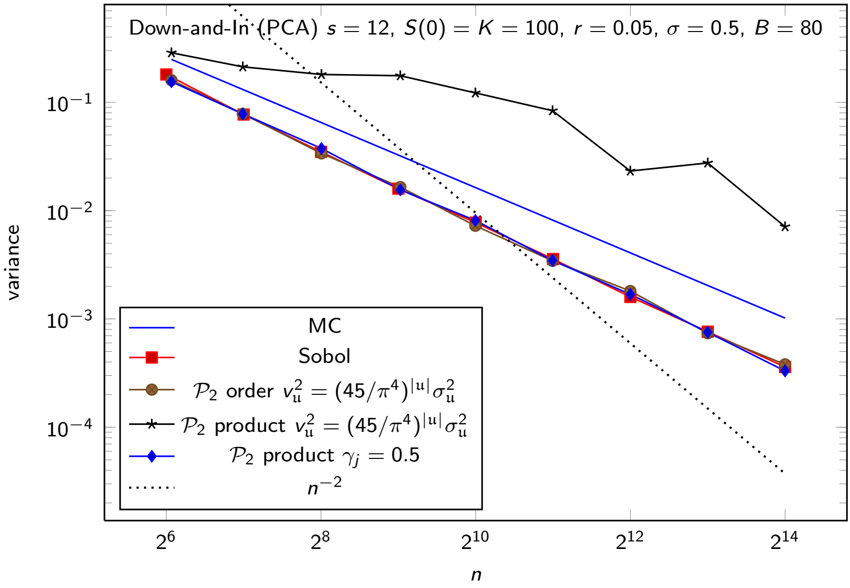
Lattices of Rank 1 with CBC



Lattices of Rank 1 with CBC



Lattices of Rank 1 with CBC

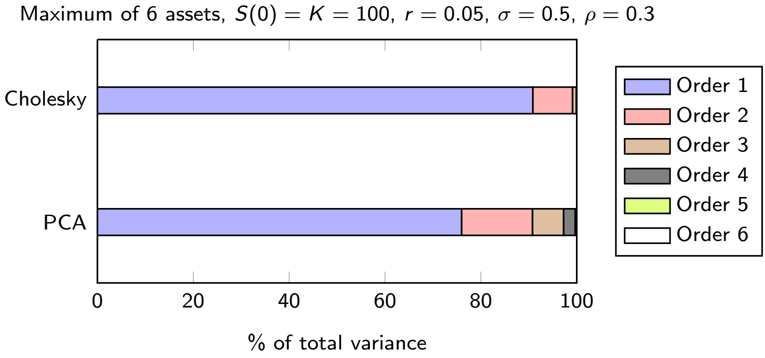


Call on the maximum of 6 assets

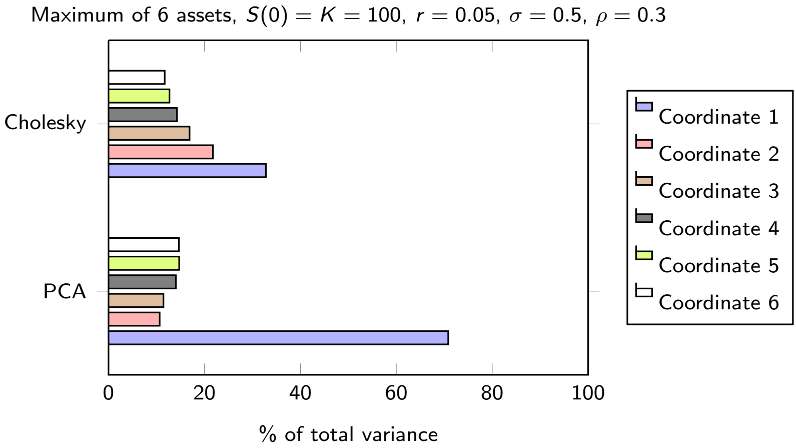
Each of 6 asset prices obeys a GBM with $s_0 = 100$, $r = 0.05$, $\sigma = 0.2$. The pairwise correlation between Brownian motions is 0.3.

The assets pay a dividend at rate 0.10, which means that the effective risk-free rate can be taken as $r' = 0.05 - 0.10 = -0.05$.

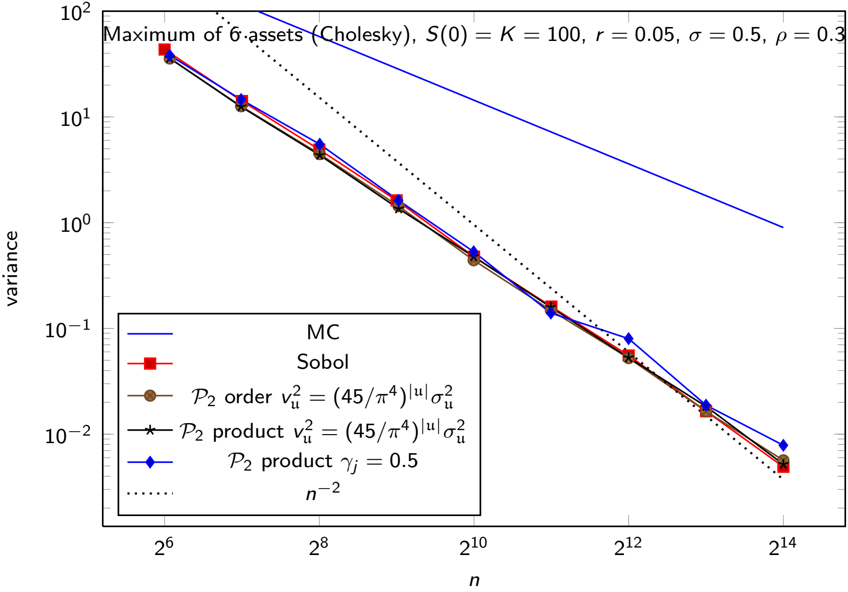
ANOVA variances for the maximum of 6 assets



Total Variance per Coordinate for max of 6 assets



Lattices of Rank 1 with CBC



Lattices of Rank 1 with CBC