Asymptotic properties of Truncated Stochastic Algorithms

Jérôme Lelong

http://cermics.enpc.fr/˜lelong

Thursday January 24th, 2008

Outline

1. The General Framework
   - A standard stochastic algorithm
   - Truncated stochastic algorithm
2. Convergence results
   - A.s convergence
   - Convergence rate
3. Application to variance reduction
   - Presentation of the problem
   - The importance sampling based strategy
   - Procedure
   - Numerical results
   - Basket Option

General Framework

Let $u \colon x \in \mathbb{R}^d \longmapsto u(x) \in \mathbb{R}^d$, be a continuous function defined as an expectation.
Assume u is untractable. We can only access u up to a measurement error, i.e. $u(x) + \varepsilon(x)$, where $\varepsilon(x)$ is a random noise.

Hypothesis 1 (convexity)

$\exists!\, x^\star \in \mathbb{R}^d, u(x^\star) = 0$ *and* $\forall x \in \mathbb{R}^d,\ x \neq x^\star,\ (x - x^\star) \cdot u(x) > 0.$

Remark: if u is the gradient of a strictly convex function, then u satisfies Hypothesis 1.

Problem: How to find the root of u ?

A standard stochastic algorithm

We define for $X_0 \in \mathbb{R}^d$

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1}. \tag{1}$$

- $(\delta M_n)_n$ measurement error, supposed to be a martingale increment.
- $\gamma_n > 0, \quad \gamma_n \searrow 0, \quad \sum \gamma_n = \infty$ and $\sum \gamma_n^2 < \infty$.

Theorem 1 (Robbins Monro)

Assume Hypothesis 1 and that

$$\forall n,\ \mathbb{E}[\|u(X_n) + \delta M_{n+1}\|^2 \,|\mathcal{F}_n] \leq K(1 + \|X_n\|^2)$$

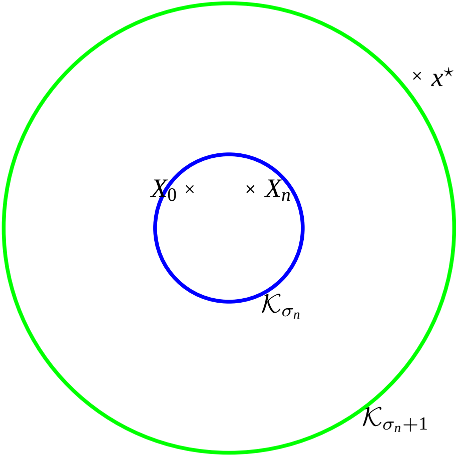
then, the sequence (1) converges a.s. to $x^\star$.

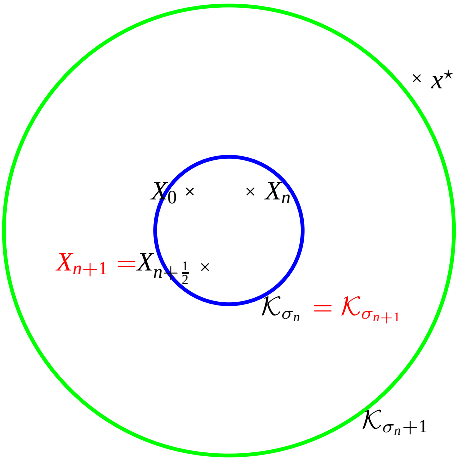
Truncated stochastic algorithm[1]: intuitive approach

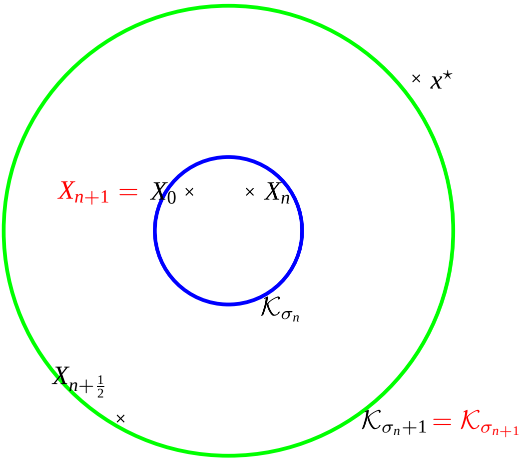
- Consider an increasing sequence of compact sets $(\mathcal{K}_j)_j$ such that $\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d$.
- Prevent the algorithm from blowing up:
  - Draw $X_{n+\frac{1}{2}}$ according to the Robbins-Monro dynamics.
  - "Project" (or truncate) it on to $\mathcal{K}_{\sigma_n}$.

  At each step, X_n should remain in a given compact set. If such is not the case, reset the algorithm and consider a larger compact set.

[1] [Chen and Zhu, 1986]


$x^\star$
X_0
X_n
$\mathcal{K}_{\sigma_n}$
$\mathcal{K}_{\sigma_n+1}$


$\times\ x^\star$
$X_0\ \times$
$\times\ X_n$
$X_{n+1} = X_{n+\frac{1}{2}}\ \times$
$\mathcal{K}_{\sigma_n} = \mathcal{K}_{\sigma_{n+1}}$
$\mathcal{K}_{\sigma_n+1}$


× $x^\star$
$X_{n+1} = X_0$ ×
× X_n
$\mathcal{K}_{\sigma_n}$
$X_{n+\frac{1}{2}}$
×
$\mathcal{K}_{\sigma_n+1} = \mathcal{K}_{\sigma_{n+1}}$

Truncated stochastic algorithm: mathematical approach

- Consider an increasing sequence of compact sets $(\mathcal{K}_j)_j$

$$\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d \quad \text{and} \quad \forall j,\ \mathcal{K}_j \subsetneq \operatorname{int}(\mathcal{K}_{j+1}) \,. \tag{2}$$

- For $X_0 \in \mathcal{K}_0$ and $\sigma_0 = 0$, we define $(X_n)_n$ and $(\sigma_n)_n$

$$\begin{cases} & X_{n+\frac{1}{2}} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1}, \\ \text{if } X_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & X_{n+1} = X_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & X_{n+1} = X_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{3}$$

σ_n counts the number of truncations up to time n.

- $\mathcal{F}_n = \sigma(X_k;\ k \leq n)$.

It is often more convenient to rewrite (3) as follows

$$\underbrace{\underbrace{X_{n+1} = X_n - \gamma_{n+1} u(X_n)}_{\text{Newton algorithm}} - \underbrace{\gamma_{n+1} \delta M_{n+1}}_{\text{noise term}}}_{\text{standard Robbins Monro algorithm}} + \underbrace{\gamma_{n+1} p_{n+1}}_{\text{truncation term}} \tag{4}$$

where

$$p_{n+1} = \begin{cases} u(X_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(X_0 - X_n) & \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}, \\ 0 & \text{otherwise.} \end{cases}$$

- δM_n is a martingale increment,
- p_n is the truncation term.

a.s convergence of truncated algorithms

Theorem 2

Assume that

- *there exists a unique* $x^\star$ *s.t.* $u(x^\star) = 0$ *and* $\forall\, x \neq x^\star, u(x) \cdot (x - x^\star) > 0$.
- $\sum_n \gamma_n = \infty$.
- *For all* $q > 0$, *the series* $\sum_n \gamma_{n+1} \delta M_{n+1} \mathbf{1}_{\{\|X_n - x^\star\| \leq q\}}$ *converges almost surely.*

Then, the sequence $(X_n)_n$ *converges a.s. to* $x^\star$ *and moreover the sequence* $(\sigma_n)_n$ *is a.s. finite (i.e. for n large enough* $p_n = 0$ *a.s.).*

a.s convergence of truncated algorithms

- $\mathbb{P}(\sigma_\infty < \infty) = 1$. But, we don't know if the sequence $(\sigma_n)_n$ is bounded.
- Assume $u(x) = \mathbb{E}(U(x, Z))$ with Z a r.v.. We can define

$$\delta M_{n+1} = U(X_n, Z_{n+1}) - u(X_n)$$

with $(Z_n)_n \overset{i.i.d.}{\sim} Z$.

$$\begin{aligned} M_n &= \sum_{i=1}^{n} \gamma_i \delta M_i \mathbf{1}_{\{\|X_{i-1}-x^\star\| \le q\}}, \\ \langle M \rangle_n &= \sum_{i=1}^{n} \gamma_i^2 \mathbb{E}(\delta M_i \delta M_i' | \mathcal{F}_{i-1}) \mathbf{1}_{\{\|X_{i-1}-x^\star\| \le q\}}, \\ \|\langle M \rangle_n\|^2 &\le \sum_{i=1}^{n} \gamma_i^2 \sup_{\|x-x^\star\| \le q} \mathbb{E}(\|U(x,Z)\|^2). \end{aligned}$$

a.s convergence of truncated algorithms

Theorem 3

Assume that

- *there exists a unique* $x^\star$ *s.t.* $u(x^\star) = 0$ *and* $\forall\, x \neq x^\star, u(x) \cdot (x - x^\star) > 0$.
- $\sum_n \gamma_n = \infty$ *and* $\sum_n \gamma_n^2 < \infty$.
- *The function* $x \longmapsto \mathbb{E}(\|U(x,Z)\|^2)$ *is bounded on any compact sets.*

Then, the sequence $(X_n)_n$ *converges a.s. to* $x^\star$ *for any sequence of compact sets satisfying (2) and moreover the sequence* $(\sigma_n)_n$ *is a.s. finite (i.e. for* n *large enough* $p_n = 0$ *a.s.).*

Averaging version

Assume that the step sequence is of the form $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$ with $1/2 < \alpha < 1$. For any $t > 0$, we introduce

$$\hat{X}_n(t) = \frac{\gamma_n}{t} \sum_{i=n}^{n+\lfloor t/\gamma_n \rfloor} X_i. \tag{5}$$

where

Theorem 4

Under the assumptions of Theorem 2, the sequence $\hat{X}_n(t)$ converges a.s. to $x^\star$ for any $t > 0$ and any sequence of compact sets satisfying (2).

Why we cannot deduce a CLT for truncated S.A. from the CLT for Standard S.A.

- The number of truncations is a.s. finite but a priori not bounded.

$$\forall \omega \in A, \mathbb{P}(A) = 1, \exists\, N(\omega), \forall n > N(\omega), p_n(\omega) = 0.$$

- N is r.v. a.s. finite but **not bounded**. Hence, one cannot use a time shift argument.
- Random time shifting does not preserve convergence in distribution.

Convergence Rate I

We are interested in the convergence of the renormalised iterates centred about their limit

$$\hat{\Delta}_n(t) = \frac{\hat{X}_n(t) - x^\star}{\sqrt{\gamma_n}}.$$

Hypothesis 2

Hypothesis on function u

- *There exists a function* $y : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *satisfying* $\lim_{\|x\|\to 0} \|y(x)\| = 0$ *and a symmetric definite positive matrix A such that*

$$u(x) = A(x - x^\star) + y(x - x^\star)(x - x^\star).$$

Convergence Rate II

Hypothesis 3

- *There exist two real numbers* $\rho > 0$ *and* $\eta > 0$ *such that*

$$\kappa = \sup_n \mathbb{E}\left(\|\delta M_n\|^{2+\rho} \mathbf{1}_{\{\|X_{n-1}-x^\star\| \leq \eta\}} \right) < \infty.$$

- *There exists a symmetric definite positive matrix* Σ *such that*

$$\mathbb{E}\left(\delta M_n \delta M_n' | \mathcal{F}_{n-1}\right) \mathbf{1}_{\{\|X_{n-1}-x^\star\| \leq \eta\}} \xrightarrow[n\to\infty]{\mathbb{P}} \Sigma.$$

Hypothesis 4

There exists $\mu > 0$ *such that* $\forall n \geq 0 \quad d(x^\star, \partial\mathcal{K}_n) > \mu.$

Convergence Rate III

Theorem 5

For $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$ *with* $1/2 < \alpha < 1$ *and under Hypotheses 1, 2, 3 and 4, the sequence* $\hat{\Delta}_n(t)$ *converges in distribution to a normally distributed random variable with mean* 0 *and variance*

$$\hat{V} = \frac{1}{t} A^{-1} \Sigma A^{-1} + \frac{A^{-2}(e^{-At} - I)V + VA^{-2}(e^{-At} - I)}{t^2}$$

where

$$V = \int_0^\infty e^{-Au} \Sigma e^{-Au} du.$$

Remarks on the hypotheses

- Hypothesis 3 is a local assumption.
  - The local integrability of $\|\delta M_n\|^{2+\rho}$ can be replaced by a condition on $\|U(x,Z)\|^{2+\rho}$ for x in a neighbourhood of $x^\star$. Some kind of Lindeberg's condition.
  - convergence of the hook of the martingale.
- Hypothesis 4 is a technical condition for the proof. In practice, no chance that $x^\star$ is on the border of one the $\mathcal{K}_n$.

Scheme of the proof

We introduce

$$\Delta_n = \frac{X_n - x^\star}{\sqrt{\gamma_n}}.$$

$\hat{\Delta}_n(t)$ can be rewritten

$$\hat{\Delta}_n(t) = \frac{\sqrt{\gamma_n}}{t} \sum_{i=n}^{n+\lfloor t/\gamma_n \rfloor} \Delta_i \sqrt{\gamma_i}.$$

Since $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$ with $\alpha < 1$, $\sqrt{\gamma_i \gamma_n} \sim \gamma_i$.
Hence,

$$\hat{\Delta}_n(t) = \frac{1}{t} \sum_{i=n}^{n+\lfloor t/\gamma_n \rfloor} \Delta_i \gamma_i.$$

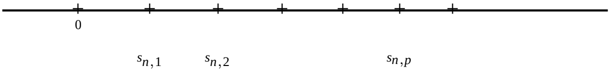
To prove Theorem 5, one needs to characterise the limiting law of $(\Delta_n, \cdots, \Delta_{n+\lfloor t/\gamma_n \rfloor})$. $\Longrightarrow$ Need of a functional result.

A functional CLT for Chen's algorithm I

For any integers k and n, we define

$$s_{n,k} = \sum_{i=n}^{n+k-1} \gamma_i \quad \text{with } s_{n,0} = 0.$$

For each fixed $n > 0$, $\mathcal{S}_n = (s_{n,k})_{k \geq 0}$ defines a discretisation grid of $[0, \infty)$ with decreasing step size γ_{n+k}.



We define $\Delta_n(\cdot)$ as the piecewise constant interpolation of $(\Delta_{n+k})_k$ on the grid $\mathcal{S}_n$. More precisely,

$$\Delta_n(0) = \Delta_n \quad \text{and} \quad \Delta_n(t) = \Delta_{n+k} \text{ if } t \in [s_{n,k}, s_{n,k+1}). \tag{6}$$

A functional CLT for Chen's algorithm II

$X_n(\cdot)$ is defined similarly. We also introduce $W_n(.)$

$$W_n(0) = 0 \quad \text{and} \quad W_n(t) = \sum_{i=n+1}^{n+k} \sqrt{\gamma_i}\delta M_i \quad \text{if } t \in [s_{n,k}, s_{n,k+1}). \tag{7}$$

Theorem 6

Assume the Hypotheses of Theorem 5.

$$(\Delta_n(\cdot), W_n(\cdot)) \xrightarrow{\mathbb{D}\times\mathbb{D}} (\Delta, W) \quad \textit{on any finite time interval}$$

where Δ is a stationary Ornstein Uhlenbeck process of initial law $\mathcal{N}(0, V)$ and W a Wiener process $\mathcal{F}^{\Delta,W}$−measurable with covariance matrix Σ.

▸ scheme of the proof

Monte Carlo computation I

We consider a diffusion Y on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$

$$\begin{cases} dY_t = Y_t r dt + \sigma(t, Y_t) dW_t, \\ S_0 = x. \end{cases} \tag{8}$$

where W is a standard Brownian motion and σ a deterministic volatility function.
We want to compute

$$P = \mathbb{E}\left(\mathrm{e}^{-rT}\,\psi(Y_t, 0 \leq t \leq T)\right).$$

Most of the time, there is no explicit solution of (8).
Let $0 = t_0 < t_1 < \cdots < t_d = T$ be a time grid, and $\hat{S}$ the discretisation of S on the grid. P can be approximated by

$$\hat{P} = \mathbb{E}\left(\mathrm{e}^{-rT}\,\psi(\hat{Y}_{t_1}, \ldots, \hat{Y}_{t_d})\right). \tag{9}$$

Monte Carlo computation II

- Using standard discretisation schemes, it is quite easy to see that $\psi(\hat{Y}_{t_1}, \dots, \hat{Y}_{t_d})$ can be expressed in terms of the Brownian increments. Equation (9) has an equivalent form

$$\hat{P} = \mathbb{E}(\phi(G)) \tag{10}$$

  where G is a $d-$dimensional Gaussian vector with identity covariance matrix.
- Expectation in (10) will be computed using Monte Carlo simulations. The challenge is to improve the convergence of the Monte Carlo procedure.
- We focus on Importance Sampling.

Importance Sampling I

An elementary change of variables in the expectation enables to change the mean of G to obtain

$$\hat{P} = \mathbb{E}\left(\phi(G+x)\, \mathrm{e}^{-x\cdot G - \frac{\|x\|^2}{2}}\right) \tag{11}$$

for any $x \in \mathbb{R}^d$.

- $\{X_x = \phi(G+x)\, \mathrm{e}^{-x\cdot G - \frac{\|x\|^2}{2}}; x \in R^d\}$ is of constant expectation.
- Find the parameter $x^\star$ that minimises $\mathrm{Var}(X_x)$.

Importance Sampling II

The variance S_2 is given by

$$S_2(x) = \mathbb{E}\left(\phi(G+x)^2\, \mathrm{e}^{-2x\cdot G - \|x\|^2}\right) = \mathbb{E}\left(\phi(G)^2\, \mathrm{e}^{-x\cdot G + \frac{\|x\|^2}{2}}\right).$$

- S_2 is strictly convex and differentiable without any particular assumptions on ψ,
- $x^\star$ is also defined as the unique root of ∇S_2

$$\nabla S_2(x) = \mathbb{E}\left((x-G)\phi(G)^2\, \mathrm{e}^{-x\cdot G + \frac{\|x\|^2}{2}}\right) = \mathbb{E}(U(x,G)).$$

Procedure

Use X_n defined by Equation (3) to approximate $x^\star$ with

$$U(x,G) = (x-G)\phi(G)^2\, \mathrm{e}^{-x\cdot G + \frac{\|x\|^2}{2}}\ .$$

Theorem 7

If there exists $\delta > 0$ such that $\mathbb{E}\,|\phi(G)|^{4+\delta} < \infty$, the sequence $X_n(t)$ defined by (5) converges a.s. to $x^\star$ for any $t > 0$.

Corollary 8

Under the hypotheses of Theorem 7, $(\hat{X}_n(t))_n$ (the averaging version of X_n) converges a.s. to $x^\star$ for any $t > 0$.

The convergence holds for any increasing sequence of compact sets $(\mathcal{K}_j)_j$.

Algorithm 1 (sequential algorithm). *Let n be the number of samples used for the Monte Carlo computation.*

1. *Draw a first set of samples following the law of G to compute an estimate of* $x^\star$*, either by using* $(X_i)_{i\leq n}$ *(see Equation (3)) or* $(\hat{X}_i)_{i\leq n}$ *(see Equation (5)). Denote the computed estimation by* $\tilde{X}$.
2. *Draw a second set of n samples following the law of G independent of the first set to compute*

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} \phi(G_i + \tilde{X})\, \mathrm{e}^{-\tilde{X}\cdot G_i - \frac{\|\tilde{X}\|^2\ T}{2}} .$$

Standard CLT for $(\Sigma_n)_n$.

Algorithm 2 (adaptive algorithm). *Let n be the number of samples used for the Monte Carlo computation.*
For each i in $0, \dots, n-1$*, do*

1. *draw a sample* G_{i+1} *according to the law of G and independent of* $\{G_j; j \leq i\}$,
2. *compute* σ_{i+1} *defined by*

$$\Sigma_{i+1} = \frac{i}{i+1}\Sigma_i + \frac{1}{i+1}\phi(G_{i+1} + X_i)\, e^{-X_i \cdot G_{i+1} - \frac{\|X_i\|^2}{2}T},$$

3. *compute* X_{i+1} *using Equation (3).*

$(\Sigma_n)_n$ satisfies a CLT for martingale arrays.

Atlas Option

Consider a basket of 16 stocks. At maturity time, remove the stocks with the three best and the three worst performances. It pays off 105% of the average of the remaining stocks.

```
basket size : 16
maturity time : 10
interest rate : 0.02
volatility : 0.2
step size : 1
Sample Number : 10000 (standard MC)
                 5000 (importance sampling)
```

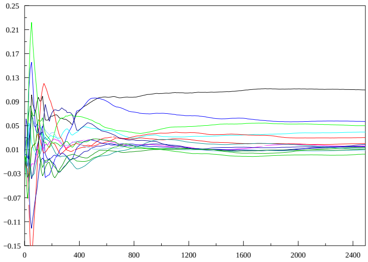


Figure: Approximation of $x^\star$ with averaging

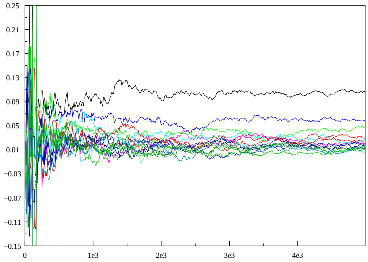


Figure: Approximation of $x^\star$ without averaging

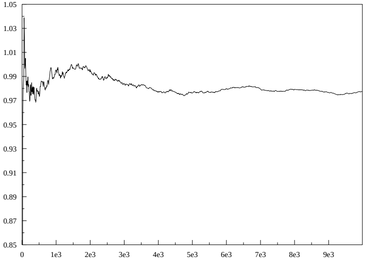


Figure: Evolution of the standard Monte Carlo simulation

variance 0.240096

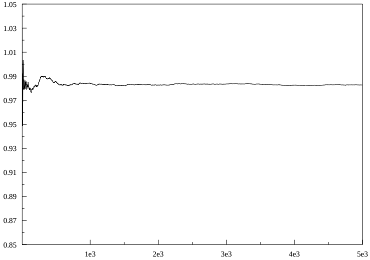


Figure: Evolution of the Monte Carlo simulation with importance sampling

variance 0.004460

Basket Option

Payoff

$$\left(\sum_{i=1}^{N} \lambda_i S_T^i - K\right)_+ .$$

```
option size 5
strike 200
initial value 50 40 60 30 55
maturity 1
volatility 0.2
interest rate   0.05
correlation 0.8
payoff coefficients 1 1 1 1 1

step size 0.01

Sample Number 5000
```

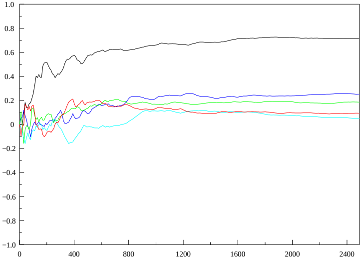


Figure: Approximation of $x^\star$ with averaging

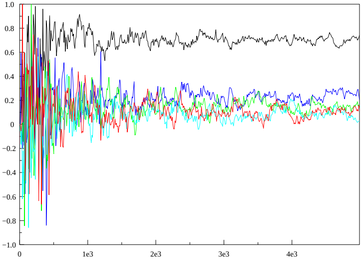


Figure: Approximation of $x^\star$ without averaging

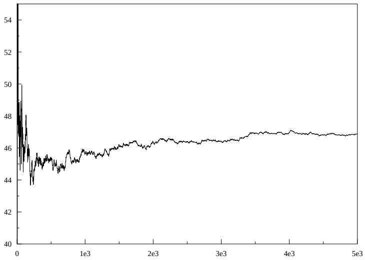


Figure: Standard Monte Carlo

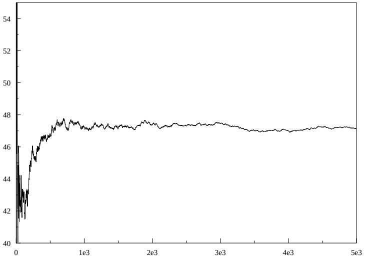


Figure: Importance Sampling coupled with MC

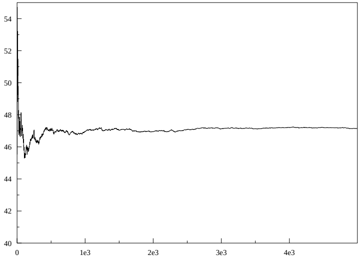


Figure: Importance Sampling + Monte Carlo (not coupled)

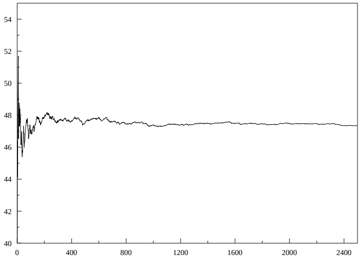


Figure: Averaging Importance Sampling with MC

The variance is divided by 10 compared with standard Monte Carlo.

Robustness of averaging algorithms

We consider the previous example and compare the convergence of the averaging and non-averaging version of the truncated algorithm for a fixed number of iterations. We extract the first component of the drift.

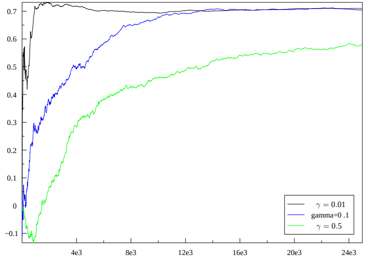


Figure: Robustness of the averaging algorithm

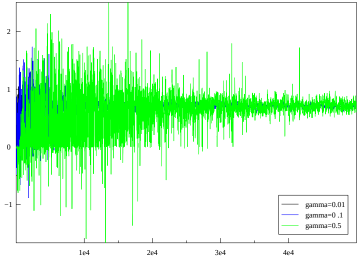


Figure: (Non) robustness of the non averaging algorithm

Limiting distribution

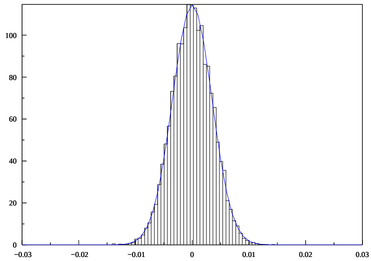


Figure: density of the averaging estimate of $x^\star$ for the Atlas option

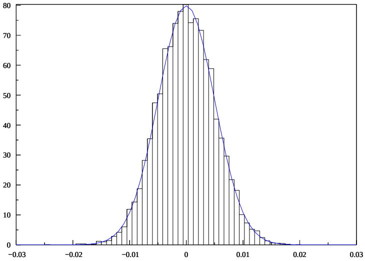


Figure: density of the non-averaging estimate of $x^\star$ for the Atlas option

major steps of the proof

- Δ_n is tight in $\mathbb{R}$. Tightness
- Use a localisation technique to prove that $(\sup_{t\in[0,T]} \|\Delta_n(t)\|)_n$ is tight.
- Prove a Donsker Theorem for martingales increment. Theorem
- $\Delta_n(\cdot)$ satisfies Aldous' criteria. Tightness in $\mathbb{D}$
- $(W_n(\cdot), \Delta_n(\cdot))_n$ is tight in $\mathbb{D} \times \mathbb{D}$ and converges in law to (W, Δ) where W is a Wiener process with respect to the smallest $\sigma-$algebra that measures $(W(\cdot), \Delta(\cdot))$ with covariance matrix Σ and Δ is the stationary solution of

$$d\Delta(t) \quad = \quad -Q\Delta(t)dt - dW(t).$$

Donsker's Theorem for martingales increments

Theorem 9

Let $(M_n(t))_n$ be a sequence of martingales. Assume that

- $(M_n(\cdot))_n$ *is tight in* $\mathbb{D}$ *and satisfies a* $\mathcal{C}-$*tightness criteria.*
- $(M_n(t))_n$ *is a uniformly square integrable family for each* t.
- $\langle M_n \rangle_t \xrightarrow[n]{\mathbb{P}} t$

Then, $M_n(\cdot) \xRightarrow[n]{\mathbb{D}[0,T]} B.M.$

$W_n(t)$ satisfies Theorem 9.

Tightness in $\mathbb{R}$

- A sequence of real valued r.v. Y_n is said to be tight if

$$\forall \varepsilon > 0,\ \exists K, \quad \mathbb{P}(|Y_n| < K) > 1 - \varepsilon \quad \forall n > 0.$$

- If such a condition holds, then from any subsequence one can extract a further subsequence which converges in distribution.
- Conversely, any weak converging sequence is tight.
- See [Billingsley, 1968] for details on convergence of probability measures.

Tightness in Skorokhod's space

- Aldous' criteria: $\forall\ \eta > 0, \quad \forall \varepsilon > 0, \quad \exists 0 < \delta < 1$ such that

$$\limsup_n \sup \left\{ \mathbb{P}\left(\|X_n(\tau) - X_n(S)\| \geq \eta \right); \begin{array}{c} S \text{ and } \tau \text{ s.t. in } [0,1], \\ S \leq \tau \leq (S+\delta) \wedge 1 \end{array} \right\} \leq \varepsilon.$$

- if $(X_n(\cdot))$ satisfies Aldous' criteria and $(\|X_n(t)\|_\infty)_n$ is tight in $\mathbb{R}$, then $X_n(\cdot)$ is tight in $\mathbb{D}[0,1]$.
- Kolmogorov's criteria ($\mathcal{C}-$ tightness) $\exists \alpha > 0, \beta > 0, \quad \forall (t,s) \in [0,1]^2$

$$\mathbb{E}(\|X_n(t) - X_n(s)\|^\alpha) \leq \kappa \left| t - s \right|^{1+\beta}.$$

if $(X_n(\cdot))$ satisfies Kolmogorov's criteria and $(\|X_n(t)\|_\infty)_n$ is tight in $\mathbb{R}$, then $(X_n(\cdot))$ is tight $\mathbb{D}[0,1]$ and any converging subsequence converges in distribution to a continuous process.

Random Time shifting

Let τ and τ' be 2 independent r.v. on $\{0,1\}$ with parameter $1/2$. We set

$$X_n := (-1)^n(\tau - \tau').$$

$\tau - \tau'$ is symmetric. Hence, X_n is constant in law.

$$\begin{aligned}
\mathbb{E}(e^{iuX_{n+\tau}}) &= \mathbb{E}(e^{iu(-1)^{n+\tau}(\tau-\tau')}),\\
&= \frac{1}{2}\left(\mathbb{E}(e^{iu(-1)^{n+\tau}\tau}) + \mathbb{E}(e^{iu(-1)^{n+\tau}(\tau-1)})\right),\\
&= \frac{1}{4}\left(1 + e^{iu(-1)^{n+1}1} + e^{iu(-1)^{n}(-1)} + 1\right),\\
&= \frac{1}{2}\left(1 + e^{iu(-1)^{n+1}}\right).
\end{aligned}$$

Hence, $(X_{n+\tau})_n$ does not converge in distribution.

Billingsley, P. (1968).
Convergence of Probability Measures.
Wiley, New York.

Chen, H. and Zhu, Y. (1986).
Stochastic Approximation Procedure with randomly varying truncations.
Scientia Sinica Series.