Robust adaptive variance reduction for normal random vectors.

Benjamin Jourdain

Joint work with Jérôme Lelong

ARC Hybrid, April 28th-29th 2009

Outline of the talk

Aim

Computation of

$$\mathbb{E}(f(G))$$

where

- $G \sim \mathcal{N}_d(0, I_d)$
- $f : \mathbb{R}^d \to \mathbb{R}$

are such that

$$\mathbb{P}(f(G) \neq 0) > 0 \text{ and } \forall \theta \in \mathbb{R}^d,\ \mathbb{E}(f^2(G)e^{-\theta.G}) < +\infty. \tag{1}$$

Let $(G_i)_{i \geq 1}$ i.i.d. $\sim \mathcal{N}_d(0, I_d)$.

Importance sampling

Let $\theta \in \mathbb{R}^d$.

$$\mathbb{E}\left(f(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}\right)=\mathbb{E}(f(G))$$

$\Rightarrow M_n(\theta,f)\stackrel{\text{def}}{=}\frac{1}{n}\sum_{i=1}^n f(G_i+\theta)e^{-\theta.G_i-\frac{|\theta|^2}{2}}$ is an a.s. convergent and asymptotically normal estimator of $\mathbb{E}(f(G))$.

$$\text{Var}(M_n(\theta,f))=\frac{1}{n}(v(\theta)-\mathbb{E}^2(f(G)))$$

where

$$v(\theta)\stackrel{\text{def}}{=}\mathbb{E}\left(f^2(G+\theta)e^{-2\theta.G-|\theta|^2}\right)=\mathbb{E}\left(f^2(G+\theta)e^{-\theta.(G+\theta)+\frac{|\theta|^2}{2}}e^{-\theta.G-\frac{|\theta|^2}{2}}\right)$$

$$\Rightarrow v(\theta)=\mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

Importance sampling

Let $\theta \in \mathbb{R}^d$.

$$\mathbb{E}\left(f(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}\right) = \mathbb{E}(f(G))$$

$\Rightarrow M_n(\theta,f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f(G_i+\theta)e^{-\theta.G_i-\frac{|\theta|^2}{2}}$ is an a.s. convergent and asymptotically normal estimator of $\mathbb{E}(f(G))$.

$$\mathrm{Var}(M_n(\theta,f)) = \frac{1}{n}(v(\theta) - \mathbb{E}^2(f(G)))$$

where

$$v(\theta) \stackrel{\text{def}}{=} \mathbb{E}\left(f^2(G+\theta)e^{-2\theta.G-|\theta|^2}\right) = \mathbb{E}\left(f^2(G+\theta)e^{-\theta.(G+\theta)+\frac{|\theta|^2}{2}}e^{-\theta.G-\frac{|\theta|^2}{2}}\right)$$

$$\Rightarrow v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

Importance sampling

Let $\theta \in \mathbb{R}^d$.

$$\mathbb{E}\left(f(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}\right) = \mathbb{E}(f(G))$$

$\Rightarrow M_n(\theta,f) \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n f(G_i+\theta)e^{-\theta.G_i-\frac{|\theta|^2}{2}}$ is an a.s. convergent and asymptotically normal estimator of $\mathbb{E}(f(G))$.

$$\text{Var}(M_n(\theta,f)) = \frac{1}{n}(v(\theta) - \mathbb{E}^2(f(G)))$$

where

$$v(\theta) \overset{\text{def}}{=} \mathbb{E}\left(f^2(G+\theta)e^{-2\theta.G-|\theta|^2}\right) = \mathbb{E}\left(f^2(G+\theta)e^{-\theta.(G+\theta)+\frac{|\theta|^2}{2}}e^{-\theta.G-\frac{|\theta|^2}{2}}\right)$$

$$\Rightarrow v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

Importance sampling

For $\theta \in \mathbb{R}^d$, $G+\theta$ admits the density $p(\theta,x) = (2\pi)^{-d/2} e^{-\frac{|x-\theta|^2}{2}}$.

$$\begin{aligned}
\mathbb{E}(f(G)) &= \int_{\mathbb{R}^d} f(x) p(0,x)\,dx \\
&= \int_{\mathbb{R}^d} f(x) \underbrace{\frac{p(0,x)}{p(\theta,x)}}_{e^{-\theta.x+\frac{|\theta|^2}{2}}} p(\theta,x)\,dx \\
&= \mathbb{E}\left(f(G+\theta) e^{-\theta.(G+\theta)+\frac{|\theta|^2}{2}} \right) \\
&= \mathbb{E}\left(f(G+\theta) e^{-\theta.G-\frac{|\theta|^2}{2}} \right)
\end{aligned}$$

Importance sampling

For $\theta \in \mathbb{R}^d$, $G+\theta$ admits the density $p(\theta,x) = (2\pi)^{-d/2} e^{-\frac{|x-\theta|^2}{2}}$.

$$\begin{aligned}
\mathbb{E}(f(G)) &= \int_{\mathbb{R}^d} f(x)p(0,x)dx \\
&= \int_{\mathbb{R}^d} f(x) \underbrace{\frac{p(0,x)}{p(\theta,x)}}_{e^{-\theta.x+\frac{|\theta|^2}{2}}} p(\theta,x)dx \\
&= \mathbb{E}\left(f(G+\theta)e^{-\theta.(G+\theta)+\frac{|\theta|^2}{2}}\right) \\
&= \mathbb{E}\left(f(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}\right)
\end{aligned}$$

Optimization of θ

Under (1) the function $v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$ is

1. C^∞ with derivatives obtained by differentiation under the expectation :

$$\nabla_\theta v(\theta) = \mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$$

$$\nabla^2_\theta v(\theta) = \mathbb{E}\left((I_d + (\theta - G)(\theta - G)^*)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

2. strongly convex.

$$\Rightarrow \exists! \theta_\star \in \mathbb{R}^d \; : \; v(\theta_\star) = \inf_{\theta \in \mathbb{R}^d} v(\theta).$$

Approximate $\mathbb{E}(f(G))$ by $M_n(\theta_\star, f)$!
Problem : v and therefore $\theta_\star$ unknown.

Optimization of θ

Under (1) the function $v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$ is

1. C^∞ with derivatives obtained by differentiation under the expectation :

$$\nabla_\theta v(\theta) = \mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$$
$$\nabla^2_\theta v(\theta) = \mathbb{E}\left((I_d + (\theta - G)(\theta - G)^*)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

2. strongly convex.

$$\Rightarrow \exists!\theta_\star \in \mathbb{R}^d \ : \ v(\theta_\star) = \inf_{\theta\in\mathbb{R}^d} v(\theta).$$

Approximate $\mathbb{E}(f(G))$ by $M_n(\theta_\star, f)$!
Problem : v and therefore $\theta_\star$ unknown.

Optimization of θ

Under (1) the function $v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$ is

1. C^∞ with derivatives obtained by differentiation under the expectation :

$$\nabla_\theta v(\theta) = \mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$$

$$\nabla^2_\theta v(\theta) = \mathbb{E}\left((I_d + (\theta - G)(\theta - G)^*)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

2. strongly convex.

$$\Rightarrow \exists! \theta_\star \in \mathbb{R}^d \;:\; v(\theta_\star) = \inf_{\theta \in \mathbb{R}^d} v(\theta).$$

Approximate $\mathbb{E}(f(G))$ by $M_n(\theta_\star, f)$!

Problem : v and therefore $\theta_\star$ unknown.

Optimization of θ

Under (1) the function $v(\theta) = \mathbb{E}\left(f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$ is

1. C^∞ with derivatives obtained by differentiation under the expectation :

$$\nabla_\theta v(\theta) = \mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)$$
$$\nabla^2_\theta v(\theta) = \mathbb{E}\left((I_d + (\theta - G)(\theta - G)^*)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right).$$

2. strongly convex.

$$\Rightarrow \exists! \theta_\star \in \mathbb{R}^d \ : \ v(\theta_\star) = \inf_{\theta\in\mathbb{R}^d} v(\theta).$$

Approximate $\mathbb{E}(f(G))$ by $M_n(\theta_\star, f)$!

Problem : v and therefore $\theta_\star$ unknown.

Optimization of θ

- *Glasserman Heidelberger Shahabuddin 99* give a large deviations argument to choose θ maximizing $\log|f(\theta)| - \frac{|\theta|^2}{2}$.
  1. only gives an approximation of $\theta_\star$,
  2. numerical search of a local maximum requires regularity of f
- *Arouna 03,04* characterizes $\theta_\star$ as the unique solution of $\mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right) = 0$ to approximate it by a Robins-Monro procedure
  1. use of the same samples to estimate $\theta_\star$ and $\mathbb{E}(f(G))$: *Arouna 04*
  2. estimator of $\mathbb{E}(f(G))$ a.s. convergent and asymptotically normal with optimal variance $v(\theta_\star) - \mathbb{E}^2(f(G))$.
  3. But need of random truncation techniques to stabilize
- *Lemaire and Pagès 08* characterize $\theta_\star$ as the unique solution of $\mathbb{E}\left((2\theta - G)f^2(G - \theta)\right) = 0$ to approximate it by a stable Robins-Monro procedure

Optimization of θ

- *Glasserman Heidelberger Shahabuddin 99* give a large deviations argument to choose θ maximizing $\log|f(\theta)| - \frac{|\theta|^2}{2}$.
  1. only gives an approximation of $\theta_\star$,
  2. numerical search of a local maximum requires regularity of f
- *Arouna 03,04* characterizes $\theta_\star$ as the unique solution of $\mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right) = 0$ to approximate it by a Robins-Monro procedure
  1. use of the same samples to estimate $\theta_\star$ and $\mathbb{E}(f(G))$: *Arouna 04*
  2. estimator of $\mathbb{E}(f(G))$ a.s. convergent and asymptotically normal with optimal variance $v(\theta_\star) - \mathbb{E}^2(f(G))$.
  3. But need of random truncation techniques to stabilize
- *Lemaire and Pagès 08* characterize $\theta_\star$ as the unique solution of $\mathbb{E}\left((2\theta - G)f^2(G-\theta)\right) = 0$ to approximate it by a stable Robins-Monro procedure

Optimization of θ

- *Glasserman Heidelberger Shahabuddin 99* give a large deviations argument to choose θ maximizing $\log|f(\theta)| - \frac{|\theta|^2}{2}$.
  1. only gives an approximation of $\theta_\star$,
  2. numerical search of a local maximum requires regularity of f
- *Arouna 03,04* characterizes $\theta_\star$ as the unique solution of $\mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right) = 0$ to approximate it by a Robins-Monro procedure
  1. use of the same samples to estimate $\theta_\star$ and $\mathbb{E}(f(G))$: *Arouna 04*
  2. estimator of $\mathbb{E}(f(G))$ a.s. convergent and asymptotically normal with optimal variance $v(\theta_\star) - \mathbb{E}^2(f(G))$.
  3. But need of random truncation techniques to stabilize
- *Lemaire and Pagès 08* characterize $\theta_\star$ as the unique solution of $\mathbb{E}\left((2\theta - G)f^2(G - \theta)\right) = 0$ to approximate it by a stable Robins-Monro procedure

Sample average optimization

Under (1), for n large enough $f(G_i) \neq 0$ for some $i \in \{1, \dots, n\}$ and the sample average approximation $v_n(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$ of v

1. C^∞ with explicit derivatives :

$$\nabla_\theta v_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(\theta - G_i)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$$

$$\nabla_\theta^2 v(\theta) = \frac{1}{n}\sum_{i=1}^{n}(I_d + (\theta - G_i)(\theta - G_i)^*)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}.$$

2. strongly convex.

$$\Rightarrow \exists!\theta_n \in \mathbb{R}^d \;:\; v_n(\theta_n) = \inf_{\theta\in\mathbb{R}^d} v_n(\theta).$$

Sample average optimization

Under (1), for n large enough $f(G_i) \neq 0$ for some $i \in \{1, \ldots, n\}$ and the sample average approximation $v_n(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$ of v

1. C^∞ with explicit derivatives :

$$\nabla_\theta v_n(\theta) = \frac{1}{n}\sum_{i=1}^n (\theta - G_i)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$$

$$\nabla^2_\theta v(\theta) = \frac{1}{n}\sum_{i=1}^n (I_d + (\theta - G_i)(\theta - G_i)^*)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}.$$

2. strongly convex.

$$\Rightarrow \exists!\theta_n \in \mathbb{R}^d \ : \ v_n(\theta_n) = \inf_{\theta\in\mathbb{R}^d} v_n(\theta).$$

Sample average optimization

Under (1), for n large enough $f(G_i) \neq 0$ for some $i \in \{1, \ldots, n\}$ and the sample average approximation $v_n(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$ of v

1. C^∞ with explicit derivatives :

$$\nabla_\theta v_n(\theta) = \frac{1}{n}\sum_{i=1}^n (\theta - G_i)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}$$

$$\nabla^2_\theta v(\theta) = \frac{1}{n}\sum_{i=1}^n (I_d + (\theta - G_i)(\theta - G_i)^*)f^2(G_i)e^{-\theta.G_i+\frac{|\theta|^2}{2}}.$$

2. strongly convex.

$$\Rightarrow \exists! \theta_n \in \mathbb{R}^d \;:\; v_n(\theta_n) = \inf_{\theta \in \mathbb{R}^d} v_n(\theta).$$

Sample average optimization

The sample approximation θ_n is characterized as the unique root of

$$\nabla_\theta v_n(\theta) = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} \Leftrightarrow \nabla_\theta u_n(\theta) = 0$$

where $u_n(\theta) \stackrel{\text{def}}{=} \frac{|\theta|^2}{2} + \log\left(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}\right)$.

$$\nabla_\theta^2 u_n(\theta) = I_d + \frac{\sum_{i=1}^n G_i G_i^* f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} - \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i} \sum_{i=1}^n G_i^* f^2(G_i) e^{-\theta.G_i}}{(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i})^2} \geq I_d.$$

$\Rightarrow$ θ_n can be computed very precisely by 4 iterations of Newton's algorithm.
Only necessitates a single computation of the payoffs $(f(G_i))_{1 \leq i \leq n}$.

Sample average optimization

The sample approximation θ_n is characterized as the unique root of

$$\nabla_\theta v_n(\theta) = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} \Leftrightarrow \nabla_\theta u_n(\theta) = 0$$

where $u_n(\theta) \stackrel{\text{def}}{=} \frac{|\theta|^2}{2} + \log\left(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}\right)$.

$$\nabla^2_\theta u_n(\theta) = I_d + \frac{\sum_{i=1}^n G_i G_i^* f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} - \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i} \sum_{i=1}^n G_i^* f^2(G_i) e^{-\theta.G_i}}{\left(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}\right)^2} \geq I_d.$$

$\Rightarrow$ θ_n can be computed very precisely by 4 iterations of Newton's algorithm.
Only necessitates a single computation of the payoffs $(f(G_i))_{1 \leq i \leq n}$.

Sample average optimization

The sample approximation θ_n is characterized as the unique root of

$$\nabla_\theta v_n(\theta) = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} \Leftrightarrow \nabla_\theta u_n(\theta) = 0$$

where $u_n(\theta) \stackrel{\text{def}}{=} \frac{|\theta|^2}{2} + \log\left(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}\right)$.

$$\nabla_\theta^2 u_n(\theta) = I_d + \frac{\sum_{i=1}^n G_i G_i^* f^2(G_i) e^{-\theta.G_i}}{\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i}} - \frac{\sum_{i=1}^n G_i f^2(G_i) e^{-\theta.G_i} \sum_{i=1}^n G_i^* f^2(G_i) e^{-\theta.G_i}}{(\sum_{i=1}^n f^2(G_i) e^{-\theta.G_i})^2} \geq I_d.$$

$\Rightarrow \theta_n$ can be computed very precisely by 4 iterations of Newton's algorithm.
Only necessitates a single computation of the payoffs $(f(G_i))_{1\leq i\leq n}$.

Robust adaptive Importance Sampling estimator

$$M_n(\theta_n, f) = \frac{1}{n}\sum_{i=1}^{n} f(G_i + \theta_n)e^{-\theta_n.G_i - \frac{|\theta_n|^2}{2}}.$$

- Use of the same samples to approximate $\theta_\star$ then $\mathbb{E}(f(G))$
- No independence between the variables $\left(f(G_i + \theta_n)e^{-\theta_n.G_i - \frac{|\theta_n|^2}{2}}\right)_{1\leq i\leq n}$

Questions :

- Convergence of the RIS estimator?
- Asymptotic normality?
- Optimal variance $v(\theta_\star) - \mathbb{E}^2(f(G))$?

Robust adaptive Importance Sampling estimator

$$M_n(\theta_n,f) = \frac{1}{n}\sum_{i=1}^{n} f(G_i+\theta_n)e^{-\theta_n.G_i-\frac{|\theta_n|^2}{2}}.$$

- Use of the same samples to approximate $\theta_\star$ then $\mathbb{E}(f(G))$
- No independence between the variables $\left(f(G_i+\theta_n)e^{-\theta_n.G_i-\frac{|\theta_n|^2}{2}}\right)_{1\leq i\leq n}$

Questions :

- Convergence of the RIS estimator?
- Asymptotic normality?
- Optimal variance $v(\theta_\star) - \mathbb{E}^2(f(G))$?

Parameter reduction

To save computation time, it may be useful to

1. introduce a matrix $A \in \mathbb{R}^{d \times d'}$ with rank $d' \leq d$,
2. approximate $\tau_\star \in \mathbb{R}^{d'}$ minimizing the strictly convex and continuous function $\mathbb{R}^{d'} \ni \tau \mapsto v(A\tau)$ by $\tau_n \in \mathbb{R}^{d'}$ minimizing the strictly convex and continuous function $\mathbb{R}^{d'} \ni \tau \mapsto v_n(A\tau)$,
3. approximate $\mathbb{E}(f(G))$ by $M_n(A\tau_n, f)$

So far, $d' = d$ and $A = I_d$.

Example : model driven by I independent Brownian motions on a time-grid $(t_1, \ldots, t_N) \to d = I \times N$.
For $d' = I$ and a good choice of A, only one change of drift parameter per Brownian motion.

Parameter reduction

To save computation time, it may be useful to

1. introduce a matrix $A \in \mathbb{R}^{d \times d'}$ with rank $d' \leq d$,
2. approximate $\tau_\star \in \mathbb{R}^{d'}$ minimizing the strictly convex and continuous function $\mathbb{R}^{d'} \ni \tau \mapsto v(A\tau)$ by $\tau_n \in \mathbb{R}^{d'}$ minimizing the strictly convex and continuous function $\mathbb{R}^{d'} \ni \tau \mapsto v_n(A\tau)$,
3. approximate $\mathbb{E}(f(G))$ by $M_n(A\tau_n, f)$

So far, $d' = d$ and $A = I_d$.

Example : model driven by I independent Brownian motions on a time-grid $(t_1, \dots, t_N) \to d = I \times N$.
For $d' = I$ and a good choice of A, only one change of drift parameter per Brownian motion.

Convergence of the importance sampling parameter

Proposition 1

1. *Under (1), τ_n and $v_n(A\tau_n)$ converge a.s. to $\tau_\star$ and $v(A\tau_\star)$.*
2. *If moreover $\forall \theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, then $\sqrt{n}(\tau_n - \tau_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$ where $B = A^*\nabla^2_\theta v(A\tau_\star)A$ and $C = \mathrm{Cov}\left(A^*(A\tau_\star - G)f^2(G)e^{-A\tau_\star\cdot G + \frac{|A\tau_\star|^2}{2}}\right)$.*

Elements of proof :

a.s. convergence of τ_n to $\tau_\star$: classical result of M-estimators

$$\sup_{|\theta|\leq M} f^2(G)e^{-\theta\cdot G + \frac{|\theta|^2}{2}} \leq e^{\frac{M^2}{2}} \quad f^2(G)\prod_{k=1}^{d}(e^{MG^k} + e^{-MG^k})$$

$\Rightarrow$ a.s. $v_n(\theta) \to v(\theta)$ locally unif. (ULLN) $\Rightarrow v_n(A\tau_n) \to v(A\tau_\star)$

Convergence of the importance sampling parameter

Proposition 1

1. *Under (1),* τ_n *and* $v_n(A\tau_n)$ *converge a.s. to* $\tau_\star$ *and* $v(A\tau_\star)$.
2. *If moreover* $\forall \theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, *then* $\sqrt{n}(\tau_n - \tau_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$ *where* $B = A^*\nabla^2_\theta v(A\tau_\star)A$ *and* $C = \mathrm{Cov}\left(A^*(A\tau_\star - G)f^2(G)e^{-A\tau_\star\cdot G + \frac{|A\tau_\star|^2}{2}}\right)$.

Elements of proof :

a.s. convergence of τ_n to $\tau_\star$: classical result of M-estimators

$$\mathbb{E}\left(\sup_{|\theta|\leq M} f^2(G)e^{-\theta\cdot G + \frac{|\theta|^2}{2}}\right) \leq e^{\frac{M^2}{2}}\mathbb{E}\left(f^2(G)\prod_{k=1}^d(e^{MG^k} + e^{-MG^k})\right) < +\infty$$

$\Rightarrow$ a.s. $v_n(\theta) \to v(\theta)$ locally unif. (ULLN) $\Rightarrow v_n(A\tau_n) \to v(A\tau_\star)$

Convergence of the importance sampling parameter

Proposition 1

1. *Under* (1), τ_n *and* $v_n(A\tau_n)$ *converge a.s. to* $\tau_\star$ *and* $v(A\tau_\star)$.
2. *If moreover* $\forall \theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, *then* $\sqrt{n}(\tau_n - \tau_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$ *where* $B = A^*\nabla^2_\theta v(A\tau_\star)A$ *and* $C = \mathrm{Cov}\left(A^*(A\tau_\star - G)f^2(G)e^{-A\tau_\star\cdot G+\frac{|A\tau_\star|^2}{2}}\right)$.

Elements of proof :

a.s. convergence of τ_n to $\tau_\star$: classical result of M-estimators

$$\mathbb{E}\left(\sup_{|\theta|\le M} f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}\right) \le e^{\frac{M^2}{2}}\,\mathbb{E}\left(f^2(G)\prod_{k=1}^{d}(e^{MG^k}+e^{-MG^k})\right) < +\infty$$

$\Rightarrow$ a.s. $v_n(\theta) \to v(\theta)$ locally unif. (ULLN) $\Rightarrow v_n(A\tau_n) \to v(A\tau_\star)$

Convergence of the importance sampling parameter

Proposition 1

1. *Under (1), τ_n and $v_n(A\tau_n)$ converge a.s. to $\tau_\star$ and $v(A\tau_\star)$.*
2. *If moreover $\forall\theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, then $\sqrt{n}(\tau_n - \tau_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$ where $B = A^*\nabla^2_\theta v(A\tau_\star)A$ and $C = \mathrm{Cov}\left(A^*(A\tau_\star - G)f^2(G)e^{-A\tau_\star\cdot G+\frac{|A\tau_\star|^2}{2}}\right)$.*

Elements of proof :

a.s. convergence of τ_n to $\tau_\star$: classical result of M-estimators

$$\mathbb{E}\left(\sup_{|\theta|\leq M} f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}\right) \leq e^{\frac{M^2}{2}}\mathbb{E}\left(f^2(G)\prod_{k=1}^{d}(e^{MG^k}+e^{-MG^k})\right) < +\infty$$

$\Rightarrow$ a.s. $v_n(\theta) \to v(\theta)$ locally unif. (ULLN) $\Rightarrow v_n(A\tau_n) \to v(A\tau_\star)$

Convergence of the importance sampling parameter

Proposition 1

1. *Under* (1), τ_n *and* $v_n(A\tau_n)$ *converge a.s. to* $\tau_\star$ *and* $v(A\tau_\star)$.
2. *If moreover* $\forall\theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, *then* $\sqrt{n}(\tau_n - \tau_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$ *where* $B = A^*\nabla^2_\theta v(A\tau_\star)A$ *and* $C = \mathrm{Cov}\left(A^*(A\tau_\star - G)f^2(G)e^{-A\tau_\star\cdot G + \frac{|A\tau_\star|^2}{2}}\right)$.

Elements of proof :

a.s. convergence of τ_n to $\tau_\star$: classical result of M-estimators

$$\mathbb{E}\left(\sup_{|\theta|\le M} f^2(G)e^{-\theta\cdot G + \frac{|\theta|^2}{2}}\right) \le e^{\frac{M^2}{2}}\mathbb{E}\left(f^2(G)\prod_{k=1}^{d}(e^{MG^k} + e^{-MG^k})\right) < +\infty$$

$\Rightarrow$ a.s. $v_n(\theta) \to v(\theta)$ locally unif. (ULLN) $\Rightarrow v_n(A\tau_n) \to v(A\tau_\star)$

Elements of proof

For $\phi(\tau,x) = f^2(x)e^{-A\tau.x+\frac{|A\tau|^2}{2}}$, $\nabla_\tau\phi(\tau,x) = A^*(A\tau - x)f^2(x)e^{-A\tau.x+\frac{|A\tau|^2}{2}}$.

$$\nabla_\tau v_n(A\tau_n) = 0 = \nabla_\tau v(A\tau_\star) \Rightarrow \frac{1}{n}\sum_{i=1}^n \nabla_\tau\phi(\tau_n, G_i) = \mathbb{E}(\nabla_\tau\phi(\tau_\star, G)).$$

Subtracting $\frac{1}{n}\sum_{i=1}^n \nabla_\tau\phi(\tau_\star, G_i)$ to both sides and multiplying by $\sqrt{n}$,

$$\underbrace{\int_0^1 \frac{1}{n}\sum_{i=1}^n \nabla^2_\tau\phi(t\tau_n + (1-t)\tau_\star, G_i)dt}_{\xrightarrow{\text{a.s.}} B} \sqrt{n}(\tau_n - \tau_\star)$$

$$= \underbrace{\sqrt{n}\left(\mathbb{E}(\nabla_\tau\phi(\tau_\star, G)) - \frac{1}{n}\sum_{i=1}^n \nabla_\tau\phi(\tau_\star, G_i)\right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0,C)}.$$

Elements of proof

For $\phi(\tau,x)=f^2(x)e^{-A\tau.x+\frac{|A\tau|^2}{2}}$, $\nabla_\tau\phi(\tau,x)=A^*(A\tau-x)f^2(x)e^{-A\tau.x+\frac{|A\tau|^2}{2}}$.

$$\nabla_\tau v_n(A\tau_n)=0=\nabla_\tau v(A\tau_\star)\Rightarrow \frac{1}{n}\sum_{i=1}^n\nabla_\tau\phi(\tau_n,G_i)=\mathbb{E}(\nabla_\tau\phi(\tau_\star,G)).$$

Subtracting $\frac{1}{n}\sum_{i=1}^n\nabla_\tau\phi(\tau_\star,G_i)$ to both sides and multiplying by $\sqrt{n}$,

$$\underbrace{\int_0^1\frac{1}{n}\sum_{i=1}^n\nabla^2_\tau\phi(t\tau_n+(1-t)\tau_\star,G_i)dt}_{\xrightarrow{\text{a.s.}}B}\sqrt{n}(\tau_n-\tau_\star)$$

$$=\underbrace{\sqrt{n}\left(\mathbb{E}(\nabla_\tau\phi(\tau_\star,G))-\frac{1}{n}\sum_{i=1}^n\nabla_\tau\phi(\tau_\star,G_i)\right)}_{\xrightarrow{\mathcal{L}}\mathcal{N}_{d'}(0,C)}.$$

Convergence of the estimator

Theorem 2

Assume (1) *and that f admits a decomposition $f = f_1 + 1_{\{d'=1\}} f_2$ with*

- f_1 *a continuous function s.t.* $\forall M > 0, \mathbb{E}\left(\sup_{|\theta| \leq M} |f_1(G+\theta)| \right) < +\infty$
- $f_2 \in \mathcal{V}_A$ *defined below.*

Then, for any deterministic integer valued sequence $(\nu_n)_n$ going to ∞ with n, $M_n(A\tau_{\nu_n}, f)$ converges a.s. to $\mathbb{E}(f(G))$.

$$\mathbb{E}\left(\sup_{|\theta| \leq M} \left(|f_1(G+\theta)| e^{-\theta \cdot G - \frac{|\theta|^2}{2}} \right) \right) \leq e^{\frac{dM^2}{2}} \sum_{\mu \in \{-M, M\}^d} \mathbb{E}\left(\sup_{|\theta| \leq M} |f_1(G+\theta+\mu)| \right)$$

$$\Rightarrow \text{ULLN : a.s. } M_n(\theta, f_1) \to \mathbb{E}(f_1(G)) \text{ locally unif.} \Rightarrow M_n(A\tau_{\nu_n}, f_1) \to \mathbb{E}(f_1(G))$$

Convergence of the estimator

Theorem 2

Assume (1) and that f admits a decomposition $f = f_1 + 1_{\{d'=1\}} f_2$ with

- *f_1 a continuous function s.t.* $\forall M > 0, \mathbb{E}\left(\sup_{|\theta| \le M} |f_1(G+\theta)| \right) < +\infty$
- $f_2 \in \mathcal{V}_A$ *defined below.*

Then, for any deterministic integer valued sequence $(\nu_n)_n$ going to ∞ with n, $M_n(A\tau_{\nu_n}, f)$ converges a.s. to $\mathbb{E}(f(G))$.

$$\sup_{|\theta|\le M}\left(|f_1(G+\theta)|e^{-\theta\cdot G-\frac{|\theta|^2}{2}}\right) \le \sup_{|\theta|\le M}|f_1(G+\theta)|\prod_{k=1}^{d}(e^{MG^k}+e^{-MG^k})$$

$$\mathbb{E}\left(\sup_{|\theta|\le M}\left(|f_1(G+\theta)|e^{-\theta\cdot G-\frac{|\theta|^2}{2}}\right)\right) \le e^{\frac{dM^2}{2}}\sum_{\mu\in\{-M,M\}^d}\mathbb{E}\left(\sup_{|\theta|\le M}|f_1(G+\theta+\mu)|\right)$$

Convergence of the estimator

Theorem 2

Assume (1) and that f admits a decomposition $f = f_1 + 1_{\{d'=1\}} f_2$ with

- f_1 *a continuous function s.t.* $\forall M > 0, \mathbb{E}\left(\sup_{|\theta| \leq M} |f_1(G+\theta)|\right) < +\infty$
- $f_2 \in \mathcal{V}_A$ *defined below.*

Then, for any deterministic integer valued sequence $(\nu_n)_n$ going to ∞ with n, $M_n(A\tau_{\nu_n}, f)$ converges a.s. to $\mathbb{E}(f(G))$.

$$\mathbb{E}\left(\sup_{|\theta|\leq M}\left(|f_1(G+\theta)|e^{-\theta\cdot G-\frac{|\theta|^2}{2}}\right)\right) \leq e^{\frac{dM^2}{2}} \sum_{\mu \in \{-M,M\}^d} \mathbb{E}\left(\sup_{|\theta|\leq M} |f_1(G+\theta+\mu)|\right)$$

$\Rightarrow$ ULLN : a.s. $M_n(\theta, f_1) \to \mathbb{E}(f_1(G))$ locally unif. $\Rightarrow M_n(A\tau_{\nu_n}, f_1) \to \mathbb{E}(f_1(G))$

Definition of $\mathcal{V}_A$

Definition 3

For $A \in \mathbb{R}^d$, we say that a function $h : \mathbb{R}^d \to \mathbb{R}$

- *is A-nondecreasing (resp. A-nonincreasing) if*

$$\forall x \in \mathbb{R}^d,\ \tau \in \mathbb{R} \mapsto h(x + A\tau) \text{ is nondecreasing (resp. nonincreasing)},$$

- *is A-monotonic if it is either A-nondecreasing or A-indecreasing,*
- *belongs to $\mathcal{V}_A$ if h may be decomposed as the sum of two A-monotonic functions g_1 and g_2 such that*

$$\exists \lambda > 0,\ \exists \beta \in [0, 2),\ \forall x \in \mathbb{R},\ |g_i(x)| \leq \lambda e^{|x|^\beta} \text{ for } i = 1, 2. \quad (2)$$

When $d = 1$, $\mathcal{V}_1$ consists of the functions of finite variation which satisfy the growth assumption (2).

Asymptotic normality

Theorem 4

Assume (1), $\forall \theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$ *and that* f *admits a decomposition* $f = f_1 + f_2 + 1_{\{d'=1\}} f_3$ *with*

1. f_1 *a* C^1 *function s.t.*
   $\forall M > 0,\ \mathbb{E}\left(\sup_{|\theta|\le M} |f_1(G+\theta)| + \sup_{|\theta|\le M} |\nabla f_1(G+\theta)|\right) < +\infty,$
2. $\exists \alpha \in \left((\sqrt{d'^2 + 8d'} - d')/4, 1\right], \beta \in [0,2), \lambda > 0,$
   $$\forall x, y \in \mathbb{R}^d, |f_2(x) - f_2(y)| \le \lambda e^{|x|^\beta \vee |y|^\beta} |x-y|^\alpha,$$
3. $f_3 \in \mathcal{V}_A$.

Then $\sqrt{n}(M_n(A\tau_n, f) - \mathbb{E}(f(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, v(A\tau_\star) - \mathbb{E}^2(f(G))\right).$

Note that $\frac{\sqrt{d'^2+8d'}-d'}{4}$ is increasing with d', equals $\frac{1}{2}$ for $d' = 1$ and converges to 1 as $d' \to \infty$.

Applications

In multidimensional Black-Scholes models or discretized local volatility models (I stocks on the time-grid $(t_1, \ldots, t_N) \rightarrow d = I \times N$), hypothesis on f_2 satisfied with $\alpha = 1$ (Lipschitz case) by

- basket Call and Put payoffs
- basket exchange payoffs
- best-of options payoffs

Problem with barrier or binary options unless $d' = 1$ and payoff in $\mathcal{V}_A$

Example : payoffs of discrete time barrier Call and Put options in the one-dimensional ($I = 1$) Black-Scholes model when $A = (\sqrt{t_1}, \sqrt{t_2 - t_1}, \ldots, \sqrt{t_N - t_{N-1}})$

Applications

In multidimensional Black-Scholes models or discretized local volatility models (I stocks on the time-grid $(t_1, \ldots, t_N) \to d = I \times N$), hypothesis on f_2 satisfied with $\alpha = 1$ (Lipschitz case) by

- basket Call and Put payoffs
- basket exchange payoffs
- best-of options payoffs

Problem with barrier or binary options unless $d' = 1$ and payoff in $\mathcal{V}_A$

Example : payoffs of discrete time barrier Call and Put options in the one-dimensional ($I = 1$) Black-Scholes model when $A = (\sqrt{t_1}, \sqrt{t_2 - t_1}, \ldots, \sqrt{t_N - t_{N-1}})$

Applications

In multidimensional Black-Scholes models or discretized local volatility models (I stocks on the time-grid $(t_1, \ldots, t_N) \to d = I \times N$), hypothesis on f_2 satisfied with $\alpha = 1$ (Lipschitz case) by

- basket Call and Put payoffs
- basket exchange payoffs
- best-of options payoffs

Problem with barrier or binary options unless $d' = 1$ and payoff in $\mathcal{V}_A$

Example : payoffs of discrete time barrier Call and Put options in the one-dimensional ($I = 1$) Black-Scholes model when $A = (\sqrt{t_1}, \sqrt{t_2 - t_1}, \ldots, \sqrt{t_N - t_{N-1}})$

Confidence intervals

Corollary 5

Under the assumptions of Theorem 4, if $\mathrm{Var}(f(G)) > 0$*, then*

$$\sqrt{\frac{n}{v_n(A\tau_n) - M_n^2(A\tau_n, f)}}\left(M_n(A\tau_n, f) - \mathbb{E}(f(G))\right) \xrightarrow{\mathcal{L}} \mathcal{N}_1(0,1).$$

Confidence Interval with asymptotic level α for $\mathbb{E}(f(G))$:

$$\left[M_n(A\tau_n, f) \pm \mathcal{N}^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{v_n(A\tau_n) - M_n^2(A\tau_n, f)}{n}}\right].$$

Confidence intervals

Corollary 5

Under the assumptions of Theorem 4, if $\mathrm{Var}(f(G)) > 0$*, then*

$$\sqrt{\frac{n}{v_n(A\tau_n) - M_n^2(A\tau_n, f)}}\left(M_n(A\tau_n, f) - \mathbb{E}(f(G))\right) \xrightarrow{\mathcal{L}} \mathcal{N}_1(0,1).$$

Confidence Interval with asymptotic level α for $\mathbb{E}(f(G))$:

$$\left[M_n(A\tau_n, f) \pm \mathcal{N}^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{v_n(A\tau_n) - M_n^2(A\tau_n, f)}{n}}\right].$$

Elements of proof

By the standard central limit theorem,

$$\sqrt{n}(M_n(A\tau_\star,f) - \mathbb{E}(f(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, v(A\tau_\star) - \mathbb{E}^2(f(G))\right).$$

It is enough to check that for $i \in \{1,2,3\}$,

$$\sqrt{n}\left[M_n(A\tau_n,f_i) - M_n(A\tau_\star,f_i)\right] \xrightarrow{\Pr} 0$$

Case $i = 1$: $\theta \mapsto M_n(\theta,f_1) = \frac{1}{n}\sum_{i=1}^n f(G_i+\theta)e^{-\theta.G_i-\frac{|\theta|^2}{2}}$ is C^1 with $\nabla_\theta M_n(\theta,f_1) = M_n(\theta,\bar{f}_1)$ where $\bar{f}_1 = \nabla f_1(x) - xf_1(x)$.
Mean value theorem $\Rightarrow \exists \bar{\tau}_n \in [\tau_n,\tau_\star]$,

$$\sqrt{n}\left[M_n(A\tau_n,f_1) - M_n(A\tau_\star,f_1)\right] = \underbrace{M_n(A\bar{\tau}_n,\bar{f}_1)}_{\substack{\xrightarrow{a.s.} \mathbb{E}(\bar{f}_1(G))=0 \\ \text{by ULLN}}} .A \underbrace{\sqrt{n}(\tau_n-\tau_\star)}_{\substack{\xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0,B^{-1}CB^{-1}) \\ \text{by Proposition 1}}}$$

Elements of proof

By the standard central limit theorem,

$$\sqrt{n}(M_n(A\tau_\star, f) - \mathbb{E}(f(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, v(A\tau_\star) - \mathbb{E}^2(f(G))\right).$$

It is enough to check that for $i \in \{1,2,3\}$,

$$\sqrt{n}\left[M_n(A\tau_n, f_i) - M_n(A\tau_\star, f_i)\right] \xrightarrow{\Pr} 0$$

Case $i = 1$: $\theta \mapsto M_n(\theta, f_1) = \frac{1}{n}\sum_{i=1}^n f(G_i+\theta)e^{-\theta.G_i - \frac{|\theta|^2}{2}}$ is C^1 with $\nabla_\theta M_n(\theta, f_1) = M_n(\theta, \bar{f}_1)$ where $\bar{f}_1 = \nabla f_1(x) - x f_1(x)$.
Mean value theorem $\Rightarrow \exists \bar{\tau}_n \in [\tau_n, \tau_\star]$,

$$\sqrt{n}\left[M_n(A\tau_n, f_1) - M_n(A\tau_\star, f_1)\right] = \underbrace{M_n(A\bar{\tau}_n, \bar{f}_1)}_{\substack{\xrightarrow{a.s.} \mathbb{E}(\bar{f}_1(G)) = 0 \\ \text{by ULLN}}} .A \underbrace{\sqrt{n}(\tau_n - \tau_\star)}_{\substack{\xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1}) \\ \text{by Proposition 1}}}$$

Elements of proof

By the standard central limit theorem,

$$\sqrt{n}(M_n(A\tau_\star, f) - \mathbb{E}(f(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, v(A\tau_\star) - \mathbb{E}^2(f(G))\right).$$

It is enough to check that for $i \in \{1, 2, 3\}$,

$$\sqrt{n}\left[M_n(A\tau_n, f_i) - M_n(A\tau_\star, f_i)\right] \xrightarrow{\Pr} 0$$

Case $i = 1$: $\theta \mapsto M_n(\theta, f_1) = \frac{1}{n}\sum_{i=1}^n f(G_i + \theta)e^{-\theta.G_i - \frac{|\theta|^2}{2}}$ is C^1 with $\nabla_\theta M_n(\theta, f_1) = M_n(\theta, \bar{f}_1)$ where $\bar{f}_1 = \nabla f_1(x) - x f_1(x)$.
Mean value theorem $\Rightarrow \exists \bar{\tau}_n \in [\tau_n, \tau_\star]$,

$$\sqrt{n}\left[M_n(A\tau_n, f_1) - M_n(A\tau_\star, f_1)\right] = \underbrace{M_n(A\bar{\tau}_n, \bar{f}_1)}_{\substack{\xrightarrow{a.s.} \mathbb{E}(\bar{f}_1(G)) = 0 \\ \text{by ULLN}}} .A \underbrace{\sqrt{n}(\tau_n - \tau_\star)}_{\substack{\xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1}) \\ \text{by Proposition 1}}}$$

Elements of proof

Case $i = 2$: For $\varepsilon > 0$, $\mathbb{P}\left(\sqrt{n}|M_n(A\tau_n, f_2) - M_n(A\tau_\star, f_2)| \geq \varepsilon\right)$ is smaller than

$$\underbrace{\mathbb{P}\left(n^\delta|\tau_n - \tau_\star| \geq 1\right)}_{\to 0 \text{ if } \delta < \frac{1}{2}} + n\underbrace{\mathbb{P}\left(|G| > \sqrt{2d\log n}\right)}_{\to 0}$$

$$\underbrace{+\mathbb{P}\left(\sup_{|\tau-\tau_\star|\leq\frac{1}{n^\delta}} |M_n(A\tau, f_2) - M_n(A\tau_\star, f_2)| \geq \frac{\varepsilon}{\sqrt{n}}, \max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n}\right)}_{\stackrel{\text{def}}{=} T_\delta}$$

$$\exists\gamma > 0,\ \max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n} \Rightarrow \forall\tau \in \bar{B}(0, |\tau_\star| + 1),$$

$$\sup_{|\tau'-\tau|\leq e^{-\gamma\log(n)^{\beta/2}} n^{-1/2\alpha}} |M_n(A\tau', f_2) - M_n(A\tau, f_2)| \leq \frac{\varepsilon}{2\sqrt{n}}.$$

Elements of proof

We cover $\bar{B}(\tau_\star, \frac{1}{n^\delta})$ with $K \le C(n^{\frac{1}{2\alpha}-\delta}e^{\gamma\log(n)^{\beta/2}})^{d'}$ balls with radius $e^{-\gamma\log(n)^{\beta/2}}n^{-1/2\alpha}$ and with centers $(\tau_k)_{1\le k\le K}$.

$$T_\delta \le \mathbb{P}\left(\max_{k\le K}|M_n(A\tau_k,f_2)-M_n(A\tau_\star,f_2)| \ge \frac{\varepsilon}{2\sqrt{n}}\right)$$

$$\le \sum_{k\le K}\frac{4n}{\varepsilon^2}\underbrace{\mathbb{E}\left((M_n(A\tau_k,f_2)-M_n(A\tau_\star,f_2))^2\right)}_{\le\frac{1}{n}\times\frac{C}{n^{2\delta\alpha}}}$$

$$\le \frac{CK}{n^{2\delta\alpha}} \le Cn^{\frac{d'}{2\alpha}-\delta(d'+2\alpha)}e^{\gamma d'\log(n)^{\beta/2}} \to 0 \text{ if } \delta > \frac{d'}{2\alpha(d'+2\alpha)}$$

Since $\alpha \in \left((\sqrt{d'^2+8d'}-d')/4, 1\right]$, one may choose $\delta \in (\frac{d'}{2\alpha(d'+2\alpha)}, \frac{1}{2})$.
Case $i=3$: Similar reasoning with monotonicity replacing Hölder continuity.

Elements of proof

We cover $\bar{B}(\tau_\star, \frac{1}{n^\delta})$ with $K \leq C(n^{\frac{1}{2\alpha}-\delta}e^{\gamma\log(n)^{\beta/2}})^{d'}$ balls with radius $e^{-\gamma\log(n)^{\beta/2}}n^{-1/2\alpha}$ and with centers $(\tau_k)_{1\leq k\leq K}$.

$$T_\delta \leq \mathbb{P}\left(\max_{k\leq K}|M_n(A\tau_k,f_2) - M_n(A\tau_\star,f_2)| \geq \frac{\varepsilon}{2\sqrt{n}}\right)$$

$$\leq \sum_{k\leq K}\frac{4n}{\varepsilon^2}\underbrace{\mathbb{E}\left((M_n(A\tau_k,f_2)-M_n(A\tau_\star,f_2))^2\right)}_{\leq\frac{1}{n}\times\frac{C}{n^{2\delta\alpha}}}$$

$$\leq \frac{CK}{n^{2\delta\alpha}} \leq Cn^{\frac{d'}{2\alpha}-\delta(d'+2\alpha)}e^{\gamma d'\log(n)^{\beta/2}} \to 0 \text{ if } \delta > \frac{d'}{2\alpha(d'+2\alpha)}$$

Since $\alpha \in \left((\sqrt{d'^2+8d'}-d')/4, 1\right]$, one may choose $\delta \in (\frac{d'}{2\alpha(d'+2\alpha)}, \frac{1}{2})$.

Case $i=3$: Similar reasoning with monotonicity replacing Hölder continuity.

Elements of proof

We cover $\bar{B}(\tau_\star, \frac{1}{n^\delta})$ with $K \leq C(n^{\frac{1}{2\alpha}-\delta} e^{\gamma \log(n)^{\beta/2}})^{d'}$ balls with radius $e^{-\gamma \log(n)^{\beta/2}} n^{-1/2\alpha}$ and with centers $(\tau_k)_{1 \leq k \leq K}$.

$$\begin{aligned} T_\delta &\leq \mathbb{P}\left(\max_{k \leq K} |M_n(A\tau_k, f_2) - M_n(A\tau_\star, f_2)| \geq \frac{\varepsilon}{2\sqrt{n}}\right) \\ &\leq \sum_{k \leq K} \frac{4n}{\varepsilon^2} \underbrace{\mathbb{E}\left((M_n(A\tau_k, f_2) - M_n(A\tau_\star, f_2))^2\right)}_{\leq \frac{1}{n} \times \frac{C}{n^{2\delta\alpha}}} \\ &\leq \frac{CK}{n^{2\delta\alpha}} \leq C n^{\frac{d'}{2\alpha} - \delta(d'+2\alpha)} e^{\gamma d' \log(n)^{\beta/2}} \to 0 \text{ if } \delta > \frac{d'}{2\alpha(d'+2\alpha)} \end{aligned}$$

Since $\alpha \in \left((\sqrt{d'^2 + 8d'} - d')/4, 1\right]$, one may choose $\delta \in (\frac{d'}{2\alpha(d'+2\alpha)}, \frac{1}{2})$.
Case $i = 3$: Similar reasoning with monotonicity replacing Hölder continuity.

Binary option in the Black-Sholes model

$$S_t = S_0 e^{\sigma W_t + (r - \frac{\sigma^2}{2})t}$$

- payoff : $1_{\{S_T \geq K\}}$
- parameters : $T = 1$, $S_0 = 100$, $\sigma = 0.2$, $r = 0.05$, $K = 140$
- Exact price : 0.05968
- For 100 000 confidence intervals with asymptotic level 95% generated by the RIS procedure according to Corollary 5, the true price falls outside 5 104 times → empirical level : 94.9%

Multidimensional Black-Scholes model

$$dS_t^i = S_t^i(rdt + \sigma^i dW_t^i),\ 1 \leq i \leq I$$

where $\langle W^i, W^j \rangle_t = (\rho \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j})t$ with $\rho \in (-\frac{1}{d-1}, 1)$.
Let L denote the lower triangular matrix involved in the Cholesky decomposition $(\rho \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j})_{1 \leq i,j \leq I} = LL^*$.
Simulation of $W = (W^1, \ldots, W^I)$ on the time-grid $0 < t_1 < t_2 < \ldots < t_N$:

$$\begin{pmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_{N-1}} \\ W_{t_N} \end{pmatrix} = \begin{pmatrix} \sqrt{t_1}L & 0 & 0 & \ldots & 0 \\ \sqrt{t_1}L & \sqrt{t_2 - t_1}L & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sqrt{t_{N-1} - t_{N-2}}L & 0 \\ \sqrt{t_1}L & \sqrt{t_2 - t_1}L & \ldots & \sqrt{t_{N-1} - t_{N-2}}L & \sqrt{t_N - t_{N-1}}L \end{pmatrix} G,$$

where $G \sim \mathcal{N}_d(0, I_d)$ with $d = I \times N$.

Basket options

Payoff : $(\sum_{i=1}^{I} \omega^i S_T^i - K)_+ \to d = I$

ρ	K	Price	Price MC	Variance MC	Price RIS	Variance RIS
0.1	45	7.210	7.216	12.12	7.209	1.04
	55	0.561	0.567	1.90	0.559	0.14
0.2	50	3.298	3.304	13.56	3.296	1.74
0.5	45	7.662	7.678	42.2	7.650	5.06
	55	1.906	1.879	14.46	1.906	1.25
0.9	45	8.215	8.154	69.47	8.211	7.89
	55	2.823	2.823	30.08	2.819	2.58

Table: Basket option in dimension $d = I = 40$ with $r = 0.05$, $T = 1$, $S_0^i = 50$, $\sigma^i = 0.2$, $\omega^i = \frac{1}{d}$ for all $i = 1, \ldots, I$ and $n = 10\,000$.

In comparison with MC, variance divided by 10 and computation time multiplied by 3 (4.5 CPU seconds instead of 1.5) $\to$ time needed to achieve a given precision divided by 3.3.

Exchange option

Payoff : $(\frac{1}{I}\sum_{i=1}^{I/2} S_T^i - \frac{1}{I}\sum_{i=I/2+1}^{I} S_T^i)_+ \rightarrow d = I$

Price	Variance MC	Variance RIS
3.58	21.66	2.97
0.129	0.511	0.016
7.4	34.04	5.02
1.08	5.24	0.52

Table: Basket option in dimension $I = 10$ with $r = 0.05$, $T = 1$, $\rho = 0.2$. The spots are chosen uniformly in $[70, 130]$ and the volatilities in $[0.1, 0.3]$. $n = 100\,000$.

- Variance divided by at least 7
- Computation time multiplied by 2
- Time needed to achieve a given precision divided by 3.5.

One-dimensional barrier option

Payoff : $(S_T - K)_+ \mathbf{1}_{\forall 1 \leq j \leq d,\ S_{t_j} \geq L}$ where $t_j = \frac{jT}{d}$

- RIS : optimization of the translation parameter $\theta \in \mathbb{R}^d$
- RRIS : optimization of $A\tau$ for $\tau \in \mathbb{R}$ with $A = (\sqrt{t_1}, \dots, \sqrt{t_d - t_{d-1}})^*$. Payoff in $\mathcal{V}_A$.

L	Price	Price MC	Var MC	Var RIS	Price RRIS	Var RRIS
70	11.445	11.472	401.51	34.10	11.454	34.33
80	11.244	11.240	401.04	35.68	11.261	36.11
90	9.689	9.672	383.93	42.54	9.705	45.37
95	7.564	7.518	342.05	42.01	7.557	49.84

Table: Down and Out Call option with $\sigma = 0.2$, $r = 0.05$, $T = 2$, $S_0^1 = 100$, $K = 110$ and $n = 10\,000$.

- Variance similar for RIS and RRIS and divided by a least 7/ MC
- Computation time multiplied by 2 for RRIS → Time needed to achieve a given precision divided by 3.5

One-dimensional barrier option

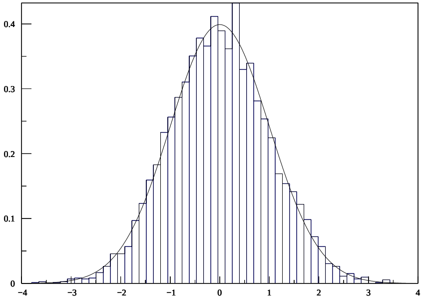

Figure: Normalized distribution of $M_n(\theta_n, f)$ (RIS) for the option with $L = 80$, $n = 10\,000$, 5 000 independent runs.

Barrier basket option

Payoff : $(\sum_{i=1}^{I} \omega^i S_T^i - K)_+ \mathbf{1}_{\forall i \leq I,\ \forall j \leq N,\ S^i_{t_j} \geq L^i}$ with $t_j = \frac{jT}{N} \rightarrow d = I \times N$.

RRIS : $d' = I$, $A_{(j-1)I+i,i} = \sqrt{t_j - t_{j-1}}$ for $j = 1, \ldots, N$ and $i = 1, \ldots, I$, all the other coefficients of A being zero.

K	Price	Price MC	Var MC	Var RIS	Price RRIS	Var RRIS
45	2.371	2.348	22.46	2.58	2.378	2.62
50	1.175	1.178	10.97	0.78	1.179	0.79
55	0.515	0.513	4.72	0.19	0.517	0.19

Table: Down and Out Call option in dimension $I = 5$ with $\sigma = 0.2$, $S_0 = (50, 40, 60, 30, 20)$, $L = (40, 30, 45, 20, 10)$, $\rho = 0.3$, $r = 0.05$, $T = 2$, $\omega = (0.2, 0.2, 0.2, 0.2, 0.2)$ and $n = 100\,000$.

Variance of RRIS similar to RIS, divided by 10 to 20/MC.
Computation time multiplied by 2.
Time needed to achieve a given precision divided by 5 to 10.

Best-of option in a local volatility model

- $dS_t^i = S_t^i(rdt + \sigma_i(t, S_t^i)dW_t^i)$ with $\sigma_i(t,x) = 0.6(1.2 - \mathrm{e}^{-0.1t}\,\mathrm{e}^{-0.001(x\,\mathrm{e}^{rt} - S_0^i)^2})\,\mathrm{e}^{-0.05\sqrt{t}}$ and $(W^1, \dots, W^I)$ like before.
- payoff : $(\max_{1 \leq i \leq I} S_T^i - K)_+$

K	Price	Var MC	Var RRIS
70	22.13	873	238
80	16.63	730	194
90	12.31	578	147

Table: Best Of option in dimension $I = 12$ with $\rho = 0.5$, $r = 0.05$, $T = 1$, $n = 50\,000$ and $S_0^i = 50$ for all $i = 1 \cdots I$.

Time needed to achieve a given precision divided by 2.

Conclusion

- Fully automatic adaptive importance sampling technique for the computation of $\mathbb{E}(f(G))$ where $f:\mathbb{R}^d\to\mathbb{R}$ and $G\sim\mathcal{N}_d(0,I_d)$.
- Theoretical results ensure convergence of the estimator and asymptotic normality with optimal limiting variance for a large class of financial payoffs f
- According to our numerical experiments,
  - time needed to achieve a given precision is divided by a factor between 2 and 10 in comparison with crude Monte Carlo
  - only one importance sampling parameter per Stock is enough
  - convergence and asymptotic normality hold for a larger class of payoffs.

    Investigation of the class of functions f s.t. $\exists\lambda>0,\beta\in[0,2)$, $\forall\varphi:\mathbb{R}^d\to\mathbb{R}^d$, C^∞ and vanishing on $B(0,M)^c$,

$$\left|\int_{\mathbb{R}^d} f\nabla.\varphi(x)dx\right|\leq\lambda e^{M^\beta}\|\varphi\|_\infty?$$

Conclusion

- Fully automatic adaptive importance sampling technique for the computation of $\mathbb{E}(f(G))$ where $f : \mathbb{R}^d \to \mathbb{R}$ and $G \sim \mathcal{N}_d(0, I_d)$.
- Theoretical results ensure convergence of the estimator and asymptotic normality with optimal limiting variance for a large class of financial payoffs f
- According to our numerical experiments,
  - time needed to achieve a given precision is divided by a factor between 2 and 10 in comparison with crude Monte Carlo
  - only one importance sampling parameter per Stock is enough
  - convergence and asymptotic normality hold for a larger class of payoffs.
    Investigation of the class of functions f s.t. $\exists \lambda > 0, \beta \in [0, 2)$, $\forall \varphi : \mathbb{R}^d \to \mathbb{R}^d$, C^∞ and vanishing on $B(0, M)^c$,

$$\left| \int_{\mathbb{R}^d} f \nabla . \varphi(x) dx \right| \leq \lambda e^{M^\beta} \|\varphi\|_\infty ?$$

Conclusion

- Fully automatic adaptive importance sampling technique for the computation of $\mathbb{E}(f(G))$ where $f : \mathbb{R}^d \to \mathbb{R}$ and $G \sim \mathcal{N}_d(0, I_d)$.
- Theoretical results ensure convergence of the estimator and asymptotic normality with optimal limiting variance for a large class of financial payoffs f
- According to our numerical experiments,
  - time needed to achieve a given precision is divided by a factor between 2 and 10 in comparison with crude Monte Carlo
  - only one importance sampling parameter per Stock is enough
  - convergence and asymptotic normality hold for a larger class of payoffs.
    Investigation of the class of functions f s.t. $\exists \lambda > 0, \beta \in [0, 2)$, $\forall \varphi : \mathbb{R}^d \to \mathbb{R}^d$, C^∞ and vanishing on $B(0, M)^c$,

$$|\int_{\mathbb{R}^d} f \nabla.\varphi(x) dx| \leq \lambda e^{M^\beta} \|\varphi\|_\infty ?$$

For $(p,\theta,\eta) \in [0,1]\times\mathbb{R}^d\times\mathbb{R}^d$ and $U \sim \mathcal{U}[0,1]$ independent of G let

$$\tilde{G} = G + 1_{\{U\le p\}}\theta + 1_{\{U>p\}}\eta.$$

$$\mathbb{E}(f(G)) = \mathbb{E}\left(\frac{f(\tilde{G})}{pe^{\theta.\tilde{G}-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.\tilde{G}-\frac{|\eta|^2}{2}}}\right)$$

$\Rightarrow M_n(p,\theta,\eta,f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \frac{f(\tilde{G}_i)}{pe^{\theta.\tilde{G}_i-\frac{|\theta|^2}{2}}+(1-p)e^{\eta.\tilde{G}_i-\frac{|\eta|^2}{2}}}$ is an a.s. convergent estimator of $\mathbb{E}(f(G))$

$$\text{Var}(M_n(p,\theta,\eta,f)) = \frac{1}{n}(v(p,\theta,\eta)-\mathbb{E}^2(f(G)))$$

where

$$v(p,\theta,\eta) \stackrel{\text{def}}{=} \mathbb{E}\left(\frac{f^2(G)}{pe^{\theta.G-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.G-\frac{|\eta|^2}{2}}}\right)$$

Problem : loss of convexity!

For $(p,\theta,\eta) \in [0,1] \times \mathbb{R}^d \times \mathbb{R}^d$ and $U \sim \mathcal{U}[0,1]$ independent of G let

$$\tilde{G} = G + 1_{\{U \leq p\}}\theta + 1_{\{U > p\}}\eta.$$

$$\mathbb{E}(f(G)) = \mathbb{E}\left(\frac{f(\tilde{G})}{pe^{\theta.\tilde{G}-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.\tilde{G}-\frac{|\eta|^2}{2}}}\right)$$

$\Rightarrow M_n(p,\theta,\eta,f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \frac{f(\tilde{G}_i)}{pe^{\theta.\tilde{G}_i-\frac{|\theta|^2}{2}}+(1-p)e^{\eta.\tilde{G}_i-\frac{|\eta|^2}{2}}}$ is an a.s. convergent estimator of $\mathbb{E}(f(G))$

$$\mathrm{Var}(M_n(p,\theta,\eta,f)) = \frac{1}{n}(v(p,\theta,\eta) - \mathbb{E}^2(f(G)))$$

where

$$v(p,\theta,\eta) \stackrel{\text{def}}{=} \mathbb{E}\left(\frac{f^2(G)}{pe^{\theta.G-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.G-\frac{|\eta|^2}{2}}}\right)$$

Problem : loss of convexity!

For $(p, \theta, \eta) \in [0,1] \times \mathbb{R}^d \times \mathbb{R}^d$ and $U \sim \mathcal{U}[0,1]$ independent of G let

$$\tilde{G} = G + 1_{\{U \leq p\}}\theta + 1_{\{U > p\}}\eta.$$

$$\mathbb{E}(f(G)) = \mathbb{E}\left(\frac{f(\tilde{G})}{pe^{\theta.\tilde{G}-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.\tilde{G}-\frac{|\eta|^2}{2}}}\right)$$

$\Rightarrow M_n(p, \theta, \eta, f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \frac{f(\tilde{G}_i)}{pe^{\theta.\tilde{G}_i-\frac{|\theta|^2}{2}}+(1-p)e^{\eta.\tilde{G}_i-\frac{|\eta|^2}{2}}}$ is an a.s. convergent estimator of $\mathbb{E}(f(G))$

$$\text{Var}(M_n(p, \theta, \eta, f)) = \frac{1}{n}(v(p, \theta, \eta) - \mathbb{E}^2(f(G)))$$

where

$$v(p, \theta, \eta) \stackrel{\text{def}}{=} \mathbb{E}\left(\frac{f^2(G)}{pe^{\theta.G-\frac{|\theta|^2}{2}} + (1-p)e^{\eta.G-\frac{|\eta|^2}{2}}}\right)$$

Problem : loss of convexity!