Robust adaptive variance reduction for normal random vectors.

Benjamin Jourdain

Collège de France, 18 march 2011

Aim

Efficient computation of

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

where

- $G \sim \mathcal{N}_d(0, I_d)$
- $f : \mathbb{R}^d \to \mathbb{R}$ is such that $\mathbb{P}(f(G) \neq 0) > 0$ and $\mathbb{E}(f^2(G)) < +\infty$.

Motivation : If $(W_t)_{t\geq 0}$ is a Brownian motion and $F : C([0,T],\mathbb{R}) \to \mathbb{R}$, then for a suitable discrete approximation $F_d : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}(F(W_t, t \leq T)) &\simeq \mathbb{E}\left(F_d\left((W_{\frac{kT}{d}} - W_{\frac{(k-1)T}{d}})_{1\leq k\leq d}\right)\right)\\ &\simeq \mathbb{E}\left(F_d\left(\sqrt{\frac{T}{d}}G\right)\right)\end{aligned}$$

Adaptive variance reduction

- For $f : \mathbb{R}^d \to \mathbb{R}$ specified, it is possible to develop efficient variance reduction techniques (control variates, importance sampling, conditioning, stratified sampling) by a fine analysis of this function
- Some banks prefer automatic variance reduction techniques which do not require such an analysis (too many new financial products)
- Adaptive variance reduction : adaptively learn the structure of $f(G)$ from the successive random drawings $(G_i)_{i\geq 1}$ i.i.d. $\sim \mathcal{N}_d(0, I_d)$ performed to approximate $\mathbb{E}(f(G)) \to$ tune the variance reduction technique.
- Robustness $\to$ to guarantee that the computation time needed to achieve a given precision is reduced.

Outline of the talk

1. Importance Sampling
   - Convergence of the importance sampling parameter
   - Convergence of the RIS estimator
   - Numerical results
2. Stratification
   - Adaptive allocation
   - Optimization of the strata

Importance sampling

Let $p(x) = (2\pi)^{-d/2}e^{-\frac{|x|^2}{2}}$ denote the density of $\mathcal{N}_d(0, I_d)$.
For $(X_i)_{i\geq 1}$ i.i.d. $\mathbb{R}^d$-valued random vectors with density $q(x)$ such that

$$dx \text{ a.e., } f(x) = 0 \Rightarrow q(x) = 0,$$

$$\mathbb{E}\left(\frac{fp}{q}(X_1)\right) = \int_{\mathbb{R}^d} \frac{fp}{q}(x)q(x)dx = \int_{\mathbb{R}^d} f(x)p(x)dx = \mathbb{E}(f(G)).$$

$\Rightarrow$ as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^n \frac{fp}{q}(X_i) \to \mathbb{E}(f(G))$ a.s. .

$$\mathrm{Var}\left(\frac{fp}{q}(X)\right) = \underbrace{\mathbb{E}\left(\left(\frac{fp}{q}\right)^2(X)\right)}_{\geq \mathbb{E}^2\left(\frac{|f|p}{q}(X)\right) = \left(\int_{\mathbb{R}^d}|f(x)|p(x)dx\right)^2} - \mathbb{E}^2(f(G))$$

with lower bound attained for $q(x) = \frac{|f|p(x)}{\mathbb{E}(|f|(G))}$ and equal to 0 is f has constant sign.

Parametric importance sampling

For $\theta \in \mathbb{R}^d$, $X_1 = G + \theta$ admits the density $p(\theta, x) = (2\pi)^{-d/2} e^{-\frac{|x-\theta|^2}{2}}$.

$$\frac{p(x)}{p(\theta, x)} = e^{-\theta . x + \frac{|\theta|^2}{2}}.$$

$$\mathbb{E}(f(G)) = \mathbb{E}\left(f(X_1)\frac{p(X_1)}{p(\theta, X_1)}\right) = \mathbb{E}\left(f(G+\theta)e^{-\theta . G - \frac{|\theta|^2}{2}}\right).$$

- Whatever the choice of θ, only necessitates the simulation of $(G_i)_{i \geq 1}$ i.i.d. $\sim \mathcal{N}_d(0, I_d)$.
- change of probability measure based on the Esscher transform

Variance for parametric importance sampling

We assume that

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

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

$\Rightarrow M_n(\theta,f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n f(G_i+\theta)e^{-\theta.G_i-\frac{|\theta|^2}{2}}$ 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

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

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

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 :

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

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

$$\geq \mathbb{E}\left(f^2(G)e^{-\frac{|G|^2}{2}}\right)I_d.$$

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 Robbins-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 Robbins-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 :

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

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

2. strongly convex as soon as $\exists i \leq n$ s.t. $f(G_i) \neq 0$.

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

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

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

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

Robust adaptive Importance Sampling estimator

Joint work with Jérôme Lelong.

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

- 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 $v(\theta_\star) - \mathbb{E}^2(f(G))$?

1. Importance Sampling
   - Convergence of the importance sampling parameter
   - Convergence of the RIS estimator
   - Numerical results

2. Stratification
   - Adaptive allocation
   - Optimization of the strata

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 $(t_1, \ldots, t_N) \rightarrow d = I \times N$.
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

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

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

1. Importance Sampling
   - Convergence of the importance sampling parameter
   - Convergence of the RIS estimator
   - Numerical results

2. Stratification
   - Adaptive allocation
   - Optimization of the strata

Convergence of the estimator

Theorem 2

Assume that $f:\mathbb{R}^d\to\mathbb{R}$ is dx a.e. continuous and such that

$$\exists\lambda>0,\ \exists\beta\in[0,2),\ \forall x\in\mathbb{R},\ |f(x)|\le\lambda e^{|x|^\beta}. \tag{2}$$

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

When f is continuous and satisfies (2), by the ULLN, a.s., $M_n(\theta,f)\to\mathbb{E}(f(G))$ locally unif. $\Rightarrow M_n(A\tau_{\nu_n},f)\to\mathbb{E}(f(G))$ Hence $\mu_n \overset{\text{def}}{=} \frac{\sum_{k=1}^n e^{-A\tau_{\nu_n}.G_k-\frac{|A\tau_{\nu_n}|^2}{2}}\delta_{G_k+A\tau_{\nu_n}}}{\sum_{k=1}^n e^{-A\tau_{\nu_n}.G_k-\frac{|A\tau_{\nu_n}|^2}{2}}} \xrightarrow{\mathcal{L}} \mathcal{N}_d(0,I_d)$ a.s.. When f is dx a.e. continuous, $\mu_n\circ f^{-1}\xrightarrow{\mathcal{L}}\mathcal{N}_d(0,I_d)\circ f^{-1}$ a.s.. Under (2), we get a.s. uniform integrability of a sequence of r.v. with laws $\mu_n\circ f^{-1}$ from the a.s. convergence of $M_n(A\tau_{\nu_n},e^{|x|^\beta})$ to $\mathbb{E}(e^{|G|^\beta})$.

Asymptotic normality

Theorem 3

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* $f = f_1 + f_2$ *with*

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

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 $d' \to \infty$.

Confidence intervals

Corollary 4

Under the assumptions of Theorem 3, 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 $\mathbb{E}(f(G))$:

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

Asymptotic normality

Remark 5

- *When $d' = 1$, a.s. convergence and asymptotic normality preserved under addition to f of $f_\downarrow + f_\uparrow$ such that*
  - $\forall x \in \mathbb{R}^d$, $\tau \in \mathbb{R} \mapsto f_\downarrow(x + A\tau)$ *is nonincreasing*; $\tau \in \mathbb{R} \mapsto f_\uparrow(x + A\tau)$ *is nondecreasing*
  - $\exists \lambda > 0,\ \exists \beta \in [0,2),\ \forall x \in \mathbb{R}^d,\ |f_\downarrow(x)| + |f_\uparrow(x)| \leq \lambda e^{|x|^\beta}$.
- *Assume that for some $k \in \mathbb{N}^*$, f is C^k with some finite moments assumptions involving its derivative up to the order k.*
  If $(\nu_n)_n$ is a deterministic sequence such that

$$\exists \lambda > 0,\ \forall n \in \mathbb{N}^*,\ \nu_n \geq \lambda n^{1/k},$$

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

1. Importance Sampling
   - Convergence of the importance sampling parameter
   - Convergence of the RIS estimator
   - Numerical results

2. Stratification
   - Adaptive allocation
   - Optimization of the strata

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 $\rho \in (-\frac{1}{d-1}, 1)$.

For $t \geq u \geq 0$, $S_t^i = S_u^i e^{\sigma^i (W_t^i - W_u^i) + (r - \frac{(\sigma^i)^2}{2})(t-u)}$.

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 $W = (W^1, \dots, W^I)$ on the time-grid $0 < t_1 < t_2 < \dots < t_N$:

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

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.

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 $A = (\sqrt{t_1}, \ldots, \sqrt{t_d - t_{d-1}})^*$. Payoff A-monotonic.

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

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

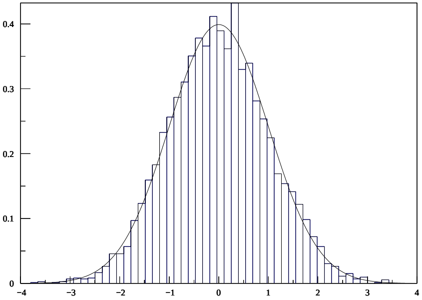

Figure: Normalized distribution of $M_n(\theta_n, f)$ (RIS) for the option with $L = 80$, $n = 10\,000$, $5\,000$ independent runs.

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 $t_j = \frac{jT}{N} \to d = I \times N$.

RRIS : $d' = I$, $A_{(j-1)I+i,i} = \sqrt{t_j - t_{j-1}}$ for $j = 1, \dots, N$ and $i = 1, \dots, I$, 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.

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
  - asymptotic normality holds 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 $B(0, M)^c$,

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

Stratification

Let $(A_i)_{1\leq i\leq I}$ be a partition of $\mathbb{R}^d$ into I strata s.t. $p_i \overset{\text{def}}{=} \mathbb{P}(G \in A_i)$ is positive and known for $i \in \{1,\dots,I\}$ and efficient simulation according to $\mathcal{L}(G|G \in A_i)$ is possible.

Example : $A_i = \left\{x \in \mathbb{R}^d :< \mu, x >\in [y_{i-1}, y_i)\right\}$ where $-\infty = y_0 < y_1 < \cdots < y_{I-1} < y_I = +\infty$ and $\mu \in \mathbb{R}^d$ is s.t. $\|\mu\| = 1$.

$p_i = \mathcal{N}(y_i) - \mathcal{N}(y_{i-1})$ where $\mathcal{N}(x) = \int_{-\infty}^x e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}$ and for $U \sim \mathcal{U}[0,1]$ indep of G,

$$G + \left[\mathcal{N}^{-1}\left((\mathcal{N}(y_{i-1}) + U(\mathcal{N}(y_i) - \mathcal{N}(y_{i-1}))\right) - < \mu, G >\right] \mu \sim \mathcal{L}(G|G \in A_i).$$

Let $(G_i^j)_{1\leq i\leq I, 1\leq j}$ be independent random variables with G_i^j distributed according to $\mathcal{L}(G|G \in A_i)$.

$$\mathbb{E}(f(G)) = \sum_{i=1}^I \mathbb{E}(f(G)|G \in A_i)\mathbb{P}(G \in A_i) = \sum_{i=1}^I p_i \mathbb{E}(f(G_i^1)).$$

Stratified estimator of $\mathbb{E}(f(G))$

Standard estimator : $\frac{1}{N}\sum_{j=1}^{N} f(G^j)$ with $(G^j)_{j\geq 1}$ i.i.d. according to the law of $G \to$ Variance :

$$v_{\text{standard}}(N) = \frac{\text{Var}(f(G))}{N} = \frac{1}{N}\left(\sum_{i=1}^{I} p_i \mathbb{E}(f^2(G_1^i)) - \underbrace{\left(\sum_{i=1}^{I} p_i \mathbb{E}(f(G_i^1))\right)^2}_{\leq \sum_{i=1}^{I} p_i \mathbb{E}^2(f(G_i^1))}\right)$$

$$\geq \frac{1}{N}\sum_{i=1}^{I} p_i \underbrace{\text{Var}(f(G_i^1))}_{\sigma_i^2}.$$

Stratified estimator : $\sum_{i=1}^{I} \frac{p_i}{N_i}\sum_{j=1}^{N_i} f(G_i^j) = \frac{1}{N}\sum_{i=1}^{I}\frac{p_i}{q_i}\sum_{j=1}^{N_i} f(G_i^j)$ where $N = \sum_{i=1}^{I} N_i$ and $q_i = \frac{N_i}{N} \to$ Variance :

$$v_{\text{stratif}}(N,q) = \frac{1}{N^2}\sum_{i=1}^{I}\frac{N_i p_i^2 \sigma_i^2}{q_i^2} = \frac{1}{N}\sum_{i=1}^{I}\frac{p_i^2\sigma_i^2}{q_i}$$

Variance reduction

Proportional allocation : $q \equiv p$ i.e. $N_i = Np_i$. Then

$$v_{\text{stratif}}(N,p) = \frac{1}{N}\sum_{i=1}^{I}\frac{p_i^2\sigma_i^2}{q_i} = \frac{1}{N}\sum_{i=1}^{I} p_i\sigma_i^2 \leq v_{\text{standard}}(N).$$

Variance reduction !

Optimal allocation :

$$Nv_{\text{stratif}}(N,q) = \sum_{i=1}^{I} q_i\left(\frac{p_i\sigma_i}{q_i}\right)^2 \geq \left(\sum_{i=1}^{I} q_i\frac{p_i\sigma_i}{q_i}\right)^2 = \left(\sum_{i=1}^{I} p_i\sigma_i\right)^2 \stackrel{\text{def}}{=} \sigma_*^2$$

with the lower bound attained for $q_i^* = \frac{p_i\sigma_i}{\sum_{l=1}^{I} p_l\sigma_l}$.

Variance even smaller but the σ_i are unknown in general.

Algorithm : joint work with Pierre Étoré

Let N^k (resp. N_i^k) denote the total number of random drawings G_i^j made in all the strata (resp. in stratum i) at the end of step k.

1. At step 1, allocate the N^1 first drawings in the strata proportionally to the p_i and estimate $\mathbb{E}(f(G_i^1))$, σ_i^2 and q_i^*,
2. At the beginning of step $k \geq 2$, allocate the $N^k - N^{k-1}$ new random drawings in the strata
   - either proportionally to the estimations $p_i\widehat{\sigma}_i^{k-1} / \sum_{l=1}^I p_l\widehat{\sigma}_l^{k-1}$ of the q_i^* available at the end of step $k-1$,
   - or in order to minimize the estimated variance $\sum_{i=1}^I (p_i\widehat{\sigma}_i^{k-1})^2/N_i^k$ of the stratified estimator after step k under the constraints $\sum_{i=1}^I N_i^k = N^k, N_i^k \geq N_i^{k-1}, \forall i \rightarrow$ explicit solution.

   Then refine the estimations of $\mathbb{E}(f(G_i^1))$, σ_i^2 and q_i^* using these new drawings.

Convertion to $\mathbb{N}_+^I$ of the above allocations which belong to $\mathbb{R}_+^I$ by some rounding procedure preserving the sum.

Forced drawings

If $\widehat{\sigma}^1_{i_0} = 0$ whereas $\sigma_{i_0} > 0$, then

- no drawings are made after step $k = 1$ in the stratum i_0.
- $\frac{1}{N^k_{i_0}} \sum_{j=1}^{N^k_{i_0}} f(G^j_{i_0}) = \frac{1}{N^1_{i_0}} \sum_{j=1}^{N^1_{i_0}} f(G^j_{i_0})$ does not converges to $\mathbb{E}(f(G^1_{i_0})) = \mathbb{E}(f(G)|G \in A_{i_0})$ when $k \to +\infty$.
- The stratified estimator $\sum_{i=1}^{I} \frac{p_i}{N^k_i} \sum_{j=1}^{N^k_i} f(G^j_i)$ does not converge to $\mathbb{E}(f(G))$.

Solution :

- choose the sequence $(N^k)_{k\geq 1}$ so that $N^k \geq N^{k-1} + I$ for all $k \geq 2$,
- enforce one drawing in each stratum at each step k,
- allocate the remaining $N^k - N^{k-1} - I$ drawings according the previous procedure.

Then

$$\forall 1 \leq i \leq I,\ \forall k \geq 1,\ \boxed{N^k_i \geq k}.$$

Convergence

Theorem 6

$$\mathbb{P}\left(\sum_{i=1}^{I}\frac{p_i}{N_i^k}\sum_{j=1}^{N_i^k} f(G_i^j) \xrightarrow[k\to\infty]{} \mathbb{E}(f(G))\right) = 1.$$

If, moreover, $\sigma_{i_0} > 0$ *for some* $i_0 \in \{1,\ldots,I\}$ *and* $\lim_{k\to+\infty}\frac{k}{N^k} = 0$*, then*

$$\sqrt{N^k}\left(\sum_{i=1}^{I}\frac{p_i}{N_i^k}\sum_{j=1}^{N_i^k} f(G_i^j) - \mathbb{E}(f(G))\right) \xrightarrow[k\to\infty]{\mathcal{L}} \mathcal{N}_1\left(0,\sigma_*^2\right)$$

with $\sigma_*^2 = \left(\sum_{i=1}^{I} p_i\sigma_i\right)^2$ *the asymptotic variance for the optimal allocation.*

$\Rightarrow \frac{\sqrt{N^k}}{\sum_{i=1}^{I} p_i\widehat{\sigma}_i^k}\left(\sum_{i=1}^{I}\frac{p_i}{N_i^k}\sum_{j=1}^{N_i^k} f(G_i^j) - \mathbb{E}(f(G))\right) \xrightarrow[k\to\infty]{\mathcal{L}} \mathcal{N}_1(0,1)$
$\rightarrow$ confidence intervals for $\mathbb{E}(f(G))$.

Adaptive optimization of the strata

Joint work with Pierre Etoré, Gersende Fort and Eric Moulines.
Assume that for $1 \leq i \leq I$, $A_i = \left\{x \in \mathbb{R}^d :< \mu, x >\in [y_{i-1}, y_i)\right\}$ where $-\infty = y_0 < y_1 < \cdots < y_{I-1} < y_I = +\infty$ and $\mu \in \mathbb{R}^d$ is s.t. $\|\mu\| = 1$. The

optimal standard deviation $\sigma_* = \sum_{i=1}^{I} p_i \sigma_i$ is equal to

$$\sum_{i=1}^{I} \sqrt{(\nu(1, y_i) - \nu(1, y_{i-1}))(\nu(f^2, y_i) - \nu(f^2, y_{i-1})) - (\nu(f, y_i) - \nu(f, y_{i-1}))^2}.$$

where $\nu(g, y) = \mathbb{E}(g(G)1_{\{<\mu,G>\leq y\}})$.

Adaptive optimization of the strata

Lemma 7

Under regularity assumptions

$$\partial_y \nu(g,y) = n(y)\mathbb{E}(g(G)| <\mu, G>= y)$$
$$\nabla_\mu \nu(g,y) = -n(y)\mathbb{E}(Gg(G)| <\mu, G>= y).$$

where $n(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$ *is the density of* $<\mu, G>$.

$$\mathbb{E}(g(G)| <\mu, G>= y) = \mathbb{E}[g(G_i^1 + (y- <\mu, G_i^1>)\mu)].$$

This enables

1. to estimate the gradient of σ_* w.r.t. $(y_1, \dots, y_{I-1})$ and μ using the random drawings G_i^j in the strata,
2. to perform a stochastic gradient descent simultaneously with the adaptive allocation algorithm.

Optimization of the boundaries

Parametrization of the boundaries by a probability density h on $\mathbb{R}$ with c.d.f. $H(y) = \int_{-\infty}^{y} h(z)dz$:

$$y_i = H^{-1}(\tfrac{i}{I}) \text{ i.e. } A_i = \left\{x \in \mathbb{R}^d :< \mu, x > \in [H^{-1}(\tfrac{i-1}{I}), H^{-1}(\tfrac{i}{I}))\right\},$$

with H^{-1} the càg pseudo-inverse of H.

Theorem 8

Assume $d \geq 2$. If for $g \in \{n, n \times \mathbb{E}(f(G)| < \mu, G >= \cdot), n \times \mathbb{E}(f^2(G)| < \mu, G >= \cdot)\}$, $\int_{\mathbb{R}} \frac{g^2}{h}(y)dy < +\infty$, *then*

$$\lim_{I \to \infty} \sigma_*(I) = \mathbb{E}\left(\sqrt{\mathrm{Var}(f(G)| < \mu, G >)}\right).$$

Limit not depending on $h \Rightarrow$ under optimal or adaptive allocation, the choice of the boundaries of the strata is not important when the number of strata is large $\rightarrow$ optimize the direction μ.

Algorithm

For $1 \leq k \leq \bar{k}$, $N_k = k \times M$.

- adaptive allocation in the strata
- initial stratification direction μ_1
- At each step k, for $g \in \{1, f, f^2\}$ and $i \in \{1, \dots, I-1\}$ compute

$$\widehat{\nabla_\mu \nu}(g, y_i)|_{\mu=\mu_k} = -\frac{n(y_i)}{N_i^k + N_{i+1}^k - N_i^{k-1} - N_{i+1}^{k-1}}$$
$$\left(\sum_{j=N_i^{k-1}+1}^{N_i^k} \tilde{g}(G_i^j + (y_i - < \mu_k, G_i^j >)\mu_k) + \sum_{j=N_{i+1}^{k-1}+1}^{N_{i+1}^k} \tilde{g}(G_{i+1}^j + (y_i - \cdots)\mu_k) \right)$$

where $\tilde{g}(x) = xg(x)$ and deduce an estimator $\widehat{\nabla_\mu \sigma_*^2}|_{\mu=\mu_k}$
Adapt the direction : $\mu_{k+1} = \mu_k - \gamma \widehat{\nabla_\mu \sigma_*^2}|_{\mu=\mu_k}$.

Numerical example : Asian option with final knockout

payoff

$$\left(\sum_{j=1}^{d} S_{\frac{jT}{d}} - K\right)_+ 1_{\{S_T \leq B\}}.$$

- $S_0 = 50, r = 0.05, T = 1, \sigma = 0.1, d = 16$
- $I = 100$ equiprobable strata given by $y_i = \mathcal{N}^{-1}\left(\frac{i}{I}\right)$
- $\bar{k} = 200, M = 20\,000$

Stratification direction

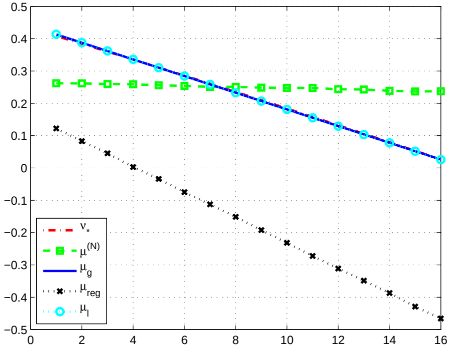

Figure: Barrier Option when $(K, B) = (50, 60)$: importance sampling parameter $\nu_* = \theta_{GHS}$

		Variance				
B	alloc	MC	AdaptStr	GHS	μ_{reg}	θ_G
60	prop	1.3393	-	0.4968	1.1466	0.4898
	adap	1.3393	0.1700	-	1.1153	0.2987
80	prop	0.70357	-	0.00107	0.00124	0.00126
	adap	0.70357	0.00046	-	0.00055	0.00057

$K = 50$, Importance sampling with θ_{GHS}
Price : 1.38 for $B = 60$ and 1.92 for $B = 80$.

Stratification along several directions

- generalization of all results to the case of stratification along several orthogonal ($\Rightarrow$ independence) directions.
- the direction $\mu_{\bar{k}}$ may be used as the first column of a rotation matrix applied to G before using Latin Hypercube Sampling
- With Bernard Lapeyre and Piergiacomo Sabino, we have developped a procedure enabling stratification of G along non-orthogonal directions.