Robust adaptive variance reduction for normal random vectors.

Jérôme Lelong (joint work with Benjamin Jourdain)

Third Conference on Numerical Methods in Finance

Wednesday 15 April 2009

Aim

Computing an expectation in a Gaussian framework

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

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

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

Outline of the talk

Importance sampling

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

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

is an a.s. convergent and asymptotically normal estimator of $\mathbb{E}(f(G))$.

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

where

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

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

Optimization of θ

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

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

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

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

2. strongly convex.

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

Approximate $\mathbb{E}(f(G))$ by $M_n(\theta^f_\star, f)$

Problem : v^f and therefore $\theta^f_\star$ 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 $\theta_\star^f$,
  2. numerical search of a local maximum requires regularity of f

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 $\theta_\star^f$,
  2. numerical search of a local maximum requires regularity of f
- *Arouna 03,04* characterizes $\theta_\star^f$ as the unique solution of $\mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right) = 0$ (stochastic approximation)
  1. use of the same samples to estimate $\theta_\star^f$ and $\mathbb{E}(f(G))$: *Arouna 04.* estimator of $\mathbb{E}(f(G))$ a.s. convergent and asymptotically normal with optimal variance $v^f(\theta_\star^f) - \mathbb{E}^2(f(G))$.
  2. But need of random truncation techniques to stabilize
- *Lemaire and Pagès 08* characterize $\theta_\star^f$ as the unique solution of $\mathbb{E}\left((2\theta - G)f^2(G - \theta)\right) = 0$ (stochastic approximation)
  $\Longrightarrow$ Tuning the gain sequence?

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

1. C^∞ with explicit derivatives :

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

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

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

1. C^∞ with explicit derivatives :

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

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

2. strongly convex.

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

Sample average optimization

The sample approximation θ_n^f is characterized as the unique root of

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

$$\begin{aligned}\nabla_\theta^2 v_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n} f^2(G_i) e^{-\theta.G_i + \frac{|\theta|^2}{2}} I_d + \text{positive semi-definite matrix} \\ &\geq \frac{1}{n}\sum_{i=1}^{n} f^2(G_i) e^{-\theta.G_i + \frac{|\theta|^2}{2}} I_d\end{aligned}$$

The lower bound can be arbitrary small $\Longrightarrow$ gradient style algorithms may converge slowly.

Sample average optimization

Alternative representation

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

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

Sample average optimization

Alternative representation

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

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

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

$\Longrightarrow \theta_n^f$ can be computed by very few steps of Newton's algorithm.

Remark : The payoffs $(f(G_i))_{1 \leq i \leq n}$ do not depend on θ.

Robust adaptive Importance Sampling estimator (RIS)

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

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

Questions :

- Convergence of the RIS estimator?
- Asymptotic normality?
- If yes, is the asymptotic variance optimal $v^f(\theta^f_\star) - \mathbb{E}^2(f(G))$?

Parameter reduction

To save computation time, it may be useful to search for the optimal θ in a sub-space $\mathcal{E} \subset \mathbb{R}^d$ with $dim(\mathcal{E}) = d' \leq d$.

1. introduce a matrix $A \in \mathbb{R}^{d \times d'}$ with rank $d' \leq d$, s.t. $Im(A) = \mathcal{E}$
2. consider $\tau_\star^{f,A} = \arg\min_{\tau \in \mathbb{R}^{d'}} v_n^f(A\tau)$ instead of $\theta_\star^f = \arg\min_{\tau \in \mathbb{R}^d} v_n^f(\theta)$. When $d' \ll d$, $\tau_\star^{f,A}$ is much easier to compute.
3. approximate $\mathbb{E}(f(G))$ by $M_n(A\tau_n^{f,A}, f)$

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

Example : model driven by a Brownian motion W in R^I on a time-grid $(t_1, \ldots, t_N) \rightarrow d = I \times N$.
The choice $W_t + \tau t$ corresponds to $d' = I$.

Convergence of the importance sampling parameter

Proposition

1. $\tau_n^{f,A}$ *and* $v_n^f(A\tau_n^{f,A})$ *converge a.s. to* $\tau_\star^{f,A}$ *and* $v^f(A\tau_\star^{f,A})$.
2. *If moreover* $\forall\theta \in \mathbb{R}^d$, $\mathbb{E}\left(f^4(G)e^{-\theta.G}\right) < +\infty$, *then*

$$\sqrt{n}(\tau_n^{f,A} - \tau_\star^{f,A}) \xrightarrow{\mathcal{L}} \mathcal{N}_{d'}(0, B^{-1}CB^{-1})$$

where $B = A^* \nabla_\theta^2 v^f(A\tau_\star^{f,A})A$ *and*

$$C = \mathrm{Cov}\left(A^*(A\tau_\star^{f,A} - G)f^2(G)e^{-A\tau_\star^{f,A}\cdot G + \frac{|A\tau_\star^{f,A}|^2}{2}}\right).$$

Definition of $\mathcal{V}_A$

Definition

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

- *is $\mathscr{A}$-monotonic if*

$$\forall x \in \mathbb{R}^d, \; \tau \in \mathbb{R} \mapsto h(x + \mathscr{A}\tau) \text{ is monotonic}$$

- *belongs to $\mathcal{V}_{\mathscr{A}}$ if h may be decomposed as the sum of two $\mathscr{A}$-monotonic functions g_1 and g_2 such that*

$$\exists \lambda > 0, \; \exists \beta \in [0,2), \; \forall x \in \mathbb{R}, \; |g_i(x)| \leq \lambda e^{|x|^\beta} \text{ for } i = 1,2. \tag{3}$$

When $d = 1$, $\mathcal{V}_1$ consists of the functions of finite variation which satisfy the growth assumption (3).

Convergence of the estimator

Theorem

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 $(\nu_n)_n \uparrow \infty$, $M_n(A\tau^{f,A}_{\nu_n}, f)$ *converges a.s. to* $\mathbb{E}(f(G))$.

Convergence of the estimator

Theorem

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 $(\nu_n)_n \uparrow \infty$, $M_n(A\tau^{f,A}_{\nu_n}, f)$ *converges a.s. to* $\mathbb{E}(f(G))$.

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

Convergence of the estimator

Theorem

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 $(\nu_n)_n \uparrow \infty$, $M_n(A\tau^{f,A}_{\nu_n}, f)$ converges a.s. to $\mathbb{E}(f(G))$.

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

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

Asymptotic normality

Theorem

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

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 $(t_1, \dots, t_N) \to d = I \times N$), hypothesis on f_2 is satisfied for

- 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 barrier Call and Put options in the one-dimensional ($I = 1$) Black-Scholes model when $A = (\sqrt{t_1}, \sqrt{t_2 - t_1}, \dots, \sqrt{t_N - t_{N-1}})$

1. Introduction
2. Convergence of the importance sampling parameter
3. Convergence of the RIS estimator
4. Numerical results

Basket options

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

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 $d = 40$ with $r = 0.05$, $T = 1$, $S_0^i = 50$, $\sigma^i = 0.2$, $\omega^i = \frac{1}{d}$ for all $i = 1, \ldots, d$ and $n = 10000$.

In comparison with MC, variance divided by 10 and computation time multiplied by 3 (4.5 CPU seconds instead of 1.5) → 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 drift 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 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

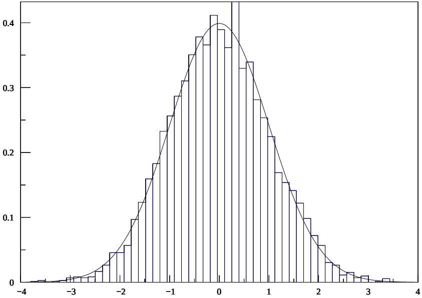

Figure: Normalized distribution of $M_n(\theta_n^f, f)$ (RIS) for the option with $L = 80$, $n = 10\,000$, $5\,000$ independent runs.

Barrier basket option

Payoff : $(\sum_{i=1}^{I} \omega^i S_T^i - K)_+ \mathbf{1}_{\{\forall i \le I, \forall j \le N, S_{t_j}^i \ge L^i\}}$ 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 $j = 1, \dots, N$ 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.

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 $\sigma_i(t,x) = 0.6(1.2 - e^{-0.1t}\, e^{-0.001(x e^{rt} - S_0^i)^2})\, e^{-0.05\sqrt{t}}$
  and $(W^1, \dots, W^I)$ as before.
- payoff : $(\max_{1 \le i \le I} S_T^i - K)_+$

K	Price	Price MC	Var MC	Price RRIS	Var RRIS
70	3.260	3.236	137	3.299	24.50
80	1.901	1.917	94.23	1.905	14.09
90	1.220	1.253	67.70	1.227	9.41

Table: Best Of option in dimension 12 with $\rho = 0.5$, $r = 0.05$, $T = 1$, $n = 50\,000$ and $\omega^i = 1$, $S_0^i = 50$ 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 $f:\mathbb{R}^d \rightarrow \mathbb{R}$ 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.