Robust Adaptive Importance Sampling for Normal Random Vectors

Benjamin Jourdain, Jérôme Lelong

Robust Adaptive Importance Sampling for Normal Random Vectors

Benjamin Jourdain[1] and Jérôme Lelong[2,3]

October 27, 2008

Abstract

Adaptive Monte Carlo methods are very efficient techniques designed to tune simulation estimators on-line. In this work, we present an alternative to stochastic approximation to tune the optimal change of measure in the context of importance sampling for normal random vectors. Unlike stochastic approximation, which requires very fine tuning in practice, we propose to use sample average approximation and deterministic optimization techniques to devise a robust and fully automatic variance reduction methodology. The same samples are used in the sample optimization of the importance sampling parameter and in the Monte Carlo computation of the expectation of interest with the optimal measure computed in the previous step. We prove that this highly non independent Monte Carlo estimator is convergent and satisfies a central limit theorem with the optimal limiting variance. Numerical experiments confirm the performance of this estimator : in comparison with the crude Monte Carlo method, the computation time needed to achieve a given precision is divided by a factor going from 3 to 15.

1 Introduction

We are interested in the computation of $\mathbb{E}(f(G))$ where $G = (G^1, \dots, G^d)$ is a d-dimensional standard normal random vector and $f : \mathbb{R}^d \to \mathbb{R}$ is a measurable function such that $f(G)$ is integrable. This problem is particularly important in mathematical finance where the calculation of the price and hedging ratios of European options in multidimensional Black-Scholes models amounts to the computation of $\mathbb{E}(f(G))$ for a well chosen function f. The same is true when the underlying assets follow more complex dynamics given by stochastic differential equations, which can be discretized using the Euler scheme for instance. We

[1]Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV, 6 et 8 avenue Blaise Pascal, 77455 Marne La Vallée, Cedex 2, France , e-mail : jourdain@cermics.enpc.fr. This research benefited from the support of the French National Research Agency (ANR) program ADAP'MC and of the "Chair Risques Financiers", Fondation du Risque

[2]Ecole Nationale Supérieure de Techniques Avancées ParisTech, Unité de Mathématiques Appliquées, 42 bd Victor 75015 Paris, e-mail : jerome.lelong@ensta.fr

[3]Projet MathFi, Université Paris-Est, INRIA Paris-Rocquencourt.

are assuming that the random variable $f(G)$ is not zero and is slightly more than square integrable :

$$\mathbb{P}(f(G) \neq 0) > 0, \tag{1.1}$$

$$\forall \theta \in \mathbb{R}^d, \mathbb{E}(f^2(G)e^{-\theta \cdot G}) < +\infty. \tag{1.2}$$

By Hölder's inequality, for $\varepsilon > 0$,

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

As a consequence, (1.2) holds as soon as

$$\exists \varepsilon > 0,\ \mathbb{E}\left(|f|^{2+\varepsilon}(G)\right) < +\infty.$$

From time to time, we will also need the following reinforced integrability condition :

$$\forall \theta \in \mathbb{R}^d,\ \mathbb{E}(f^4(G)e^{-\theta \cdot G}) < +\infty. \tag{1.3}$$

For any measurable function $h : \mathbb{R}^d \to \mathbb{R}$ either nonnegative or such that $\mathbb{E}|h(G)| < +\infty$, one has

$$\forall \theta \in \mathbb{R}^d,\ \mathbb{E}(h(G)) = \mathbb{E}\left(h(G+\theta)e^{-\theta \cdot G - \frac{|\theta|^2}{2}}\right). \tag{1.4}$$

Applying this equality to $h(x) = f(x)$ and to $h(x) = f^2(x)e^{-\theta.x+\frac{|\theta|^2}{2}}$, one obtains that the expectation and the variance of the random variable $f(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}$ are respectively equal to $\mathbb{E}(f(G))$ and $v^f(\theta) - \mathbb{E}^2(f(G))$ where

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

As a consequence, if $(G_i)_{i\geq 1}$ denotes a sequence of i.i.d. d-dimensional standard normal random vectors, for any importance sampling parameter $\theta \in \mathbb{R}^d$,

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

is an unbiased and convergent estimator of $\mathbb{E}(f(G))$. Since $n\,\mathrm{Var}(M_n(\theta, f)) = v^f(\theta) - \mathbb{E}^2(f(G))$, to improve the accuracy of the estimation for a fixed number n of random samples, one should choose θ minimizing $v^f(\theta)$. The first section of this paper addresses this minimization problem. First, we check that v^f is a strongly convex function going to infinity at infinity, which ensures the existence of a unique value $\theta^f_\star$ such that $v^f(\theta^f_\star) = \inf_{\theta \in \mathbb{R}^d} v^f(\theta)$. Of course, when $\mathbb{E}(f(G))$ is unknown, so is generally the function v^f. Therefore, direct optimization of this function is not implementable. Under (1.2), the function v^f is infinitely continuously differentiable and such that

$$\nabla_\theta v^f(\theta) = \mathbb{E}\left((\theta - G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right). \tag{1.5}$$

At this stage, we can see the interest of performing the change of measure (1.4) to transform $\mathbb{E}(f^2(G+\theta)e^{-2\theta\cdot G-|\theta|^2})$ into the above expression of v^f : no smoothness assumption on the function f is required in order to differentiate within the expectation.

Arouna [Arouna, 0304] [Arouna, 2004] takes advantage of the characterization of the optimal parameter $\theta_\star^f$ as the unique solution of the equation $\mathbb{E}\left((\theta-G)f^2(G)e^{-\theta.G+\frac{|\theta|^2}{2}}\right)=0$, in order to approximate it by a Robbins-Monro procedure. The standard Robbins-Monro algorithm explodes but it can be stabilized using random truncation techniques, see for instance [Chen et al., 1988, Chen and Zhu, 1986] or [Lelong, 2008]. According to [Arouna, 2004], the same random drawings G_i may be used to estimate simultaneously the optimal parameter $\theta_\star^f$ and the expectation of interest $\mathbb{E}(f(G))$. Moreover, both estimators are strongly consistent and the one of $\mathbb{E}(f(G))$ is asymptotically normal with an asymptotic variance equal to the optimal one $v^f(\theta_\star^f)-\mathbb{E}^2(f(G))$. Asymptotic normality of the estimator of $\theta_\star^f$ is discussed in [Lelong, 2007].

Tuning the increasing sequence of compact subsets used in randomly truncated procedures is not easy. In [Lemaire and Pagès, 2008], Lemaire and Pagès notice that using (1.4) in (1.5) leads to $\nabla_\theta v^f(\theta)=e^{|\theta|^2}\mathbb{E}((2\theta-G)f^2(G-\theta))$ and propose to use the characterization of $\theta_\star$ as the unique solution of $\mathbb{E}((2\theta-G)f^2(G-\theta))=0$ to approximate it by a Robbins-Monro procedure. As soon as the function f satisfies some exponential growth assumptions at infinity, the algorithm they propose is stable without resorting to random truncation techniques. Starting from the present Gaussian framework, Lemaire and Pagès [Lemaire and Pagès, 2008] extend this construction of non-exploding Robbins-Monro algorithms to a large class of families of multidimensional probability distributions and even to diffusion process distributions.

In the present paper, we propose and study an alternative approach, which does not require the delicate tuning of the gain sequence which is still necessary to ensure the stability of Robbins-Monro procedures. When $f(G_i)\neq 0$ for some index $i\in\{1,\dots,n\}$ (by (1.1), a.s. this condition is satisfied for n large enough), the Monte-Carlo approximation $v_n^f(\theta)\overset{\text{def}}{=}\frac{1}{n}\sum_{i=1}^n f^2(G_i)e^{-\theta\cdot G_i+\frac{|\theta|^2}{2}}$ of the function v^f is also strongly convex and going to infinity at infinity. This ensures the existence of a unique parameter θ_n^f such that $v_n^f(\theta_n^f)=\inf_{\theta\in\mathbb{R}^d}v_n^f(\theta)$. The function v_n^f is of class C^∞ and its gradient and Hessian matrices

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

are easily computed if the random samples $(G_i)_{1\leq i\leq n}$ are stored in the computer memory. Therefore, θ_n^f can be obtained by Newton's optimization procedure and we propose to estimate $\mathbb{E}(f(G))$ by $M_n(\theta_n^f,f)$. In the context of control variate variance reduction techniques, Kim and Henderson [Kim and Henderson, 2007] also propose sample average optimization of the control variate parameters as an alternative to stochastic approximation techniques. But in their algorithm, the expectation of interest is only computed in a second step involving random variables independent from the ones used in the optimization step. In our algorithm, in order to save computation time, the respective approximations θ_n^f and $M_n(\theta_n^f,f)$ of the

optimal parameters and of the expectation of interest are computed using the same random samples $(G_i)_{1\leq i\leq n}$. This makes the mathematical analysis of the properties of $M_n(\theta_n^f, f)$ more complicated : for instance, in general, $M_n(\theta_n^f, f)$ is a biased estimator of $\mathbb{E}(f(G))$. The ideea of using the same samples both for the optimization procedure and the Monte Carlo computation has mainly been investigated in the context of linear control variates. Lavenberg, Moeller and Welsh [Lavenberg et al., 1982] and Nelson [Nelson, 1990] have already noticed that for linear control variates, using the same samples does not bring any bias in the Monte Carlo estimator. Kim and Henderson [Kim and Henderson, 2004] and Glasserman [Glasserman, 2004] have also deeply investigated the linear case.

The first section of the paper is devoted to the convergence of θ_n^f to $\theta_\star^f$: almost sure convergence holds and a central limit theorem can be derived under the reinforced integrability condition (1.3). Moreover, $v_n^f(\theta_n^f)$ converges a.s. to $v^f(\theta_\star)$. The second section addresses the asymptotic properties as $n \to \infty$ of our estimator $M_n(\theta_n^f, f)$ of the expectation of interest $\mathbb{E}(f(G))$. We prove that when f is continuous and such that $\forall M > 0$, $\mathbb{E}\left(\sup_{|\theta|\leq M} |f(G+\theta)|\right) < +\infty$, then $M_n(\theta_n^f, f)$ converges a.s. to $\mathbb{E}(f(G))$. In dimension $d = 1$, this continuity assumption may be relaxed : the strong consistency of $M_n(\theta_n^f, f)$ still holds as soon as f is the sum of a continuous function as before and a function of finite variation satisfying some natural growth condition. When f satisfies (1.3) and can be decomposed as the sum of a locally Lipschitz continuous function with some natural control of the growth of the Lipschitz constant and of a C^1 function satisfying some integrability conditions, then the estimator $M_n(\theta_n^f, f)$ is asymptotically normal with optimal asymptotic variance $v^f(\theta_\star) - \mathbb{E}^2(f(G))$: $\sqrt{n}(M_n(\theta_n^f, f) - \mathbb{E}(f(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, v^f(\theta_\star) - \mathbb{E}^2(f(G))\right)$. Moreover, $\sqrt{n}\frac{M_n(\theta_n^f,f)-\mathbb{E}(f(G))}{\sqrt{v_n^f(\theta_n^f)-M_n^2(\theta_n^f,f)}} \xrightarrow{\mathcal{L}} \mathcal{N}_1(0,1)$, which enables us to construct confidence intervals for $\mathbb{E}(f(G))$. Again, in dimension $d = 1$, the conclusion is preserved if one adds a function with finite variation satisfying some natural growth condition to the previous decomposition. The third section of the paper deals with a generalization of the framework which permits to recover results concerning the weighted Monte Carlo calibration technique introduced in [Avellaneda et al., 2001] and studied in [Jourdain and Nguyen, 2001, Nguyen, 2003]. In the last section, we illustrate our theoretical results with numerical experiments which confirm the performance of our algorithm.

2 Convergence of the importance sampling parameters

According to our numerical experiments, it may be optimal, in terms of the computation time needed to achieve a given precision for the estimation of $\mathbb{E}(f(G))$, to look for the best importance sampling parameter θ in a subspace $\{A\vartheta : \vartheta \in \mathbb{R}^{d'}\}$ of $\mathbb{R}^d$ where $A \in \mathbb{R}^{d\times d'}$ is a matrix with rank $d' \leq d$. When $f(G)$ corresponds to the payoff of an option written on a d'-dimensional Black-Scholes model monitored on a regular time grid, it is sensible to use the same parameter for each coordinate G^k corresponding to a time increment of a given stock. That is why we introduce

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

Since $v^{f,A}(\vartheta) = v^f(A\vartheta)$, the properties of the function $v^{f,A}$ may be deduced from the ones of v^f. The case tackled in the introduction corresponds to the particular choice $d' = d$ and $A = I_d$.

Lemma 2.1 *Under* (1.2)*, the function* v^f *is infinitely continuously differentiable with* $\forall \alpha = (\alpha^1, \ldots, \alpha^d) \in \mathbb{N}^d$, $\forall \theta = (\theta^1, \ldots, \theta^d) \in \mathbb{R}^d$,

$$\frac{\partial^{\alpha^1+\ldots+\alpha^d}}{\partial_{\theta^1}^{\alpha^1} \ldots \partial_{\theta^d}^{\alpha^d}} v^f(\theta) = \mathbb{E}\left(\frac{\partial^{\alpha^1+\ldots+\alpha^d}}{\partial_{\theta^1}^{\alpha^1} \ldots \partial_{\theta^d}^{\alpha^d}} \left[f^2(G) e^{-\theta.G+\frac{|\theta|^2}{2}} \right] \right).$$

Under (1.1)*, the function* v^f *is strongly convex and hence such that* $\lim_{|\theta|\to+\infty} v^f(\theta) = +\infty$.

Proof : The function $\theta \mapsto f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}$ is infinitely continuously differentiable with $\frac{\partial}{\partial_{\theta^j}} f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}} = f^2(G)(\theta^j - G^j)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}$. Since,

$$\sup_{|\theta|\le M} |\partial_{\theta^j} f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}| \le e^{\frac{M^2}{2}} f^2(G) \left(M + (e^{G^j} + e^{-G^j}) \right) \prod_{k=1}^{d} (e^{MG^k} + e^{-MG^k}) \qquad (2.1)$$

where the right-hand-side is integrable by (1.2), Lebesgue's theorem ensures that v^f is continuously differentiable with $\frac{\partial}{\partial_{\theta^j}} v^f(\theta) = \mathbb{E}\left(f^2(G)(\theta^j - G^j)e^{-\theta\cdot G+\frac{|\theta|^2}{2}} \right)$. Higher order differentiability properties are obtained by similar arguments and in particular $\frac{\partial^2}{\partial_{\theta^j}\partial_{\theta^i}} v^f(\theta) = \mathbb{E}\left((\mathbf{1}_{\{i=j\}} + (\theta^j - G^j)(\theta^i - G^i)) f^2(G) e^{-\theta\cdot G+\frac{|\theta|^2}{2}} \right)$.
Assumption (1.1) ensures the existence of $\varepsilon > 0$ such that $\mathbb{P}(f^2(G) \ge \varepsilon, |G| \le \frac{1}{\varepsilon}) > 0$. Since $\mathbb{E}(f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}) \ge \varepsilon e^{-\frac{|\theta|}{\varepsilon}+\frac{|\theta|^2}{2}} \mathbb{P}(f^2(G) \ge \varepsilon, |G| \le \frac{1}{\varepsilon})$, one easily deduces that $\lim_{|\theta|\to+\infty} \mathbb{E}(f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}) = +\infty$. As the continuous function $\theta \mapsto \mathbb{E}(f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}})$ does not vanish, the Hessian matrix $\nabla^2_\theta v^f(\theta)$ is uniformly bounded from below by the positive definite matrix $\inf_{\theta\in\mathbb{R}^d} \mathbb{E}(f^2(G)e^{-\theta\cdot G+\frac{|\theta|^2}{2}}) I_d$. This yields the strong convexity of the function v^f. ■

As a consequence, $v^{f,A}$ is a strongly convex function going to infinity at infinity and there exists a unique $\vartheta^{f,A}_\star \in \mathbb{R}^{d'}$ such that $v^{f,A}(\vartheta^{f,A}_\star) = \inf_{\vartheta\in\mathbb{R}^{d'}} v^{f,A}(\vartheta)$.

Let $(G_i)_{i\ge1}$ be a sequence of d-dimensional independent standard normal random variables. For n large enough, $f(G_i) \neq 0$ for some index $i \in \{1, \ldots, n\}$ and the approximation $v^{f,A}_n(\vartheta) = \frac{1}{n}\sum_{i=1}^n f^2(G_i)e^{-A\vartheta.G_i+\frac{|A\vartheta|^2}{2}}$ of $v^{f,A}(\vartheta)$ is also strongly convex and such that $\lim_{|\vartheta|\to+\infty} v^{f,A}_n(\vartheta) = +\infty$. Hence, there exists a unique $\vartheta^{f,A}_n \in \mathbb{R}^{d'}$ such that $v^{f,A}_n(\vartheta^{f,A}_n) = \inf_{\vartheta\in\mathbb{R}^{d'}} v^{f,A}_n(\vartheta)$. The following proposition describes the asymptotic behavior of $\vartheta^{f,A}_n$ as $n \to \infty$.

Proposition 2.2 *Under* (1.1) *and* (1.2)*,* $\vartheta^{f,A}_n$ *and* $v^{f,A}_n(\vartheta^{f,A}_n)$ *converge a.s. to* $\vartheta^{f,A}_\star$ *and* $v^{f,A}(\vartheta^{f,A}_\star)$ *as* $n \to \infty$*. If, moreover,* (1.3) *holds, then* $\sqrt{n}(\vartheta^{f,A}_n - \vartheta^{f,A}_\star) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Gamma)$ *where*

$$\Gamma = [\nabla^2 v^{f,A}(\vartheta^{f,A}_\star)]^{-1} \mathrm{Cov}\left(A^*(A\vartheta^{f,A}_\star - G) f^2(G) e^{-A\vartheta^{f,A}_\star\cdot G+\frac{|A\vartheta^{f,A}_\star|^2}{2}} \right) [\nabla^2 v^{f,A}(\vartheta^{f,A}_\star)]^{-1},$$

and $\nabla^2 v^{f,A}(\vartheta) = \mathbb{E}\left(A^*(I_d + (A\vartheta - G)(A\vartheta - G)^*)Af^2(G)e^{-A\vartheta\cdot G + \frac{|A\vartheta|^2}{2}}\right)$.

Remark 2.3 *The Hessian matrix* $\nabla^2 v^{f,A}(\vartheta)$ *is positive definite under* (1.1) *and* (1.2). *If, moreover,* (1.3) *holds, using the inequality* $|G^k| \leq \mathrm{e}^{G^k} + \mathrm{e}^{-G^k}$ *for all* $1 \leq k \leq d$, *one obtains that the covariance matrix* $\mathrm{Cov}\left((A\vartheta_\star^{f,A} - G)f^2(G)e^{-A\vartheta_\star^{f,A}\cdot G + \frac{|A\vartheta_\star^{f,A}|^2}{2}}\right)$ *exists and* Γ *is well defined.*
To get some insights on the expression of this asymptotic covariance matrix, notice that if $\phi(\vartheta, x) = f^2(x)e^{-A\vartheta.x + \frac{|A\vartheta|^2}{2}}$, *subtracting* $\frac{1}{n}\sum_{i=1}^n \nabla_\vartheta \phi(\vartheta_\star^{f,A}, G_i)$ *to both sides of the equality* $\nabla v_n^{f,A}(\vartheta_n^{f,A}) = \nabla v^{f,A}(\vartheta_\star^{f,A})$ *and multiplying by* $\sqrt{n}$, *one obtains*

$$\int_0^1 \frac{1}{n}\sum_{i=1}^n \nabla_\theta^2 \phi(t\vartheta_n^{f,A} + (1-t)\vartheta_\star^{f,A}, G_i)dt\sqrt{n}(\vartheta_n^{f,A} - \vartheta_\star^{f,A})$$
$$= \sqrt{n}\left(\mathbb{E}(\nabla_\vartheta \phi(\vartheta_\star^{f,A}, G)) - \frac{1}{n}\sum_{i=1}^n \nabla_\vartheta \phi(\vartheta_\star^{f,A}, G_i)\right).$$

To prove the Proposition, we use the following uniform strong law of large numbers, which is a restatement of [Rubinstein and Shapiro, 1993, Lemma A1]. This result is also a consequence of the strong law of large numbers in Banach spaces [Ledoux and Talagrand, 1991, Corollary 7.10, page 189].

Lemma 2.4 *Let* $(X_i)_{i\geq 1}$ *be a sequence of i.i.d.* $\mathbb{R}^m$*-valued random vectors and* $h : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ *be a measurable function. Assume that*

- *a.s.,* $\theta \in \mathbb{R}^d \mapsto h(\theta, X_1)$ *is continuous,*
- $\forall M > 0,\ \mathbb{E}\left(\sup_{|\theta|\leq M} |h(\theta, X_1)|\right) < +\infty$.

Then, a.s. $\theta \in \mathbb{R}^d \mapsto \frac{1}{n}\sum_{i=1}^n h(\theta, X_i)$ *converges locally uniformly to the continuous function* $\theta \in \mathbb{R}^d \mapsto \mathbb{E}(h(\theta, X_1))$.

Proof of Proposition 2.2 : Since for $M > 0$,

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

where the right-hand-side is integrable by (1.2), applying Lemma 2.4 with $(X_i)_{i\geq 1} = (G_i)_{i\geq 1}$ and $h(\theta, x) = f^2(x)e^{-\theta.x + \frac{|\theta|^2}{2}}$ ensures that a.s., v_n^f converges locally uniformly to v^f. We restrict ourselves to a subset with probability one of the original probability space on which this convergence holds. Let $\varepsilon > 0$. By the strict convexity and the continuity of $v^{f,A}$,

$$\alpha \stackrel{\text{def}}{=} \inf_{\vartheta : |\vartheta - \vartheta_\star^{f,A}| = \varepsilon} v^{f,A}(\vartheta) - v^{f,A}(\vartheta_\star^{f,A}) > 0.$$

The local uniform convergence of $v_n^{f,A}$ to $v^{f,A}$ ensures that

$$\exists n_\alpha \in \mathbb{N}^*,\ \forall n \geq n_\alpha,\ \forall \vartheta \ s.t.\ |\vartheta - \vartheta_\star^{f,A}| \leq \varepsilon,\ |v_n^{f,A}(\vartheta) - v^{f,A}(\vartheta)| \leq \frac{\alpha}{3}.$$

For $n \geq n_\alpha$ and ϑ such that $|\vartheta - \vartheta_\star^{f,A}| \geq \varepsilon$, we deduce, using the convexity of $v_n^{f,A}$ for the first inequality,

$$\begin{aligned} v_n^{f,A}(\vartheta) - v_n^{f,A}(\vartheta_\star^{f,A}) &\geq \frac{|\vartheta - \vartheta_\star^{f,A}|}{\varepsilon}\left[v_n^{f,A}\left(\vartheta_\star^{f,A} + \varepsilon\frac{\vartheta - \vartheta_\star^{f,A}}{|\vartheta - \vartheta_\star^{f,A}|}\right) - v_n^{f,A}(\vartheta_\star^{f,A})\right] \\ &\geq \frac{|\vartheta - \vartheta_\star^{f,A}|}{\varepsilon}\left[v^{f,A}\left(\vartheta_\star^{f,A} + \varepsilon\frac{\vartheta - \vartheta_\star^{f,A}}{|\vartheta - \vartheta_\star^{f,A}|}\right) - v^{f,A}(\vartheta_\star^{f,A}) - \frac{2\alpha}{3}\right] \geq \frac{\alpha}{3}. \end{aligned}$$

Since $v_n^{f,A}(\vartheta_n^{f,A}) \leq v_n^{f,A}(\vartheta_\star^{f,A})$, we conclude that $|\vartheta_n^{f,A} - \vartheta_\star^{f,A}| < \varepsilon$ for $n \geq n_\alpha$. Therefore, $\vartheta_n^{f,A}$ converges a.s. to $\vartheta_\star^{f,A}$. By combining this last result with the local uniform convergence of $v_n^{f,A}$ to the continuous function $v^{f,A}$, we deduce that $v_n^{f,A}(\vartheta_n^{f,A})$ converges a.s. to $v^{f,A}(\vartheta_\star^{f,A})$.

By (2.1) and (1.2), for $M > 0$, $\mathbb{E}\left(\sup_{|\theta|\leq M} |\nabla_\theta f^2(G) e^{-\theta\cdot G + \frac{|\theta|^2}{2}}|\right) < +\infty$.

Similarly, $\mathbb{E}\left(\sup_{|\theta|\leq M} |\nabla_\theta^2 f^2(G) e^{-\theta\cdot G + \frac{|\theta|^2}{2}}|\right) < +\infty$. The central limit theorem governing the convergence of $\vartheta_n^{f,A}$ to $\vartheta_\star^{f,A}$ ensues from [Rubinstein and Shapiro, 1993, TheoremA2]. ■

3 Strong Law of Large Numbers and Central Limit Theorem

Let

$$\theta_n^{f,A} = A\vartheta_n^{f,A} \quad \text{and} \quad \theta_\star^{f,A} = A\vartheta_\star^{f,A}.$$

The convergence of our estimator $M_n(\theta_n^{f,A}, f)$ of $\mathbb{E}(f(G))$ is ensured by the following theorem, which is a consequence of Propositions 3.6 and 3.13 below. As we do not take advantage of the definition of $\theta_n^{f,A}$ but only use its convergence properties obtained in the previous section, these propositions deal with the asymptotic properties of $M_n(\theta_n^{f,A}, g)$ where $g : \mathbb{R}^d \to \mathbb{R}$ is an arbitrary function and

$$\forall \theta \in \mathbb{R}^d,\ M_n(\theta, g) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n g(G_i + \theta) e^{-\theta\cdot G_i - \frac{|\theta|^2}{2}}.$$

To precise the hypotheses on f in the case $d' = 1$ of a one-dimensional importance sampling parameter ϑ, we introduce the following definition.

Definition 3.1 *For $\mathcal{A} \in \mathbb{R}^d$, we say that a function $h : \mathbb{R}^d \to \mathbb{R}$*

- *is $\mathcal{A}$-nondecreasing (resp. $\mathcal{A}$-nonincreasing) if*

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

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

$$\exists \lambda > 0,\ \exists \beta \in [0,2),\ \forall x \in \mathbb{R},\ |g_i(x)| \leq \lambda e^{|x|^\beta} \text{ for } i = 1, 2. \tag{3.1}$$

When $d = 1$, $\mathcal{V}_1$ consists of the functions of finite variation which satisfy the growth assumption (3.1).

Theorem 3.2 *Assume* (1.1), (1.2) *and that f admits a decomposition $f = f_1 + \mathbf{1}_{\{d'=1\}} f_2$ with f_1 a continuous function such that $\forall M > 0$, $\mathbb{E}\left(\sup_{|\theta| \leq M} |f_1(G+\theta)|\right) < +\infty$ and $f_2 \in \mathcal{V}_A$. Then, for any deterministic integer valued sequence $(\nu_n)_n$ going to ∞ with n, $M_n(\theta_{\nu_n}^{f,A}, f)$ converges a.s. to $\mathbb{E}(f(G))$.*

Note that for the integrability condition on f_1 to hold, it is enough that $\exists \beta \in [0,2), \lambda > 0$, $\forall x \in \mathbb{R}^d$, $|f_1(x)| \leq \lambda e^{|x|^\beta}$.
Under stronger assumptions on f, the convergence of $M_n(\theta_n^{f,A}, f)$ to $\mathbb{E}(f(G))$ is governed by a central limit theorem with optimal asymptotic variance $v^{f,A}(\vartheta_\star^{f,A}) - \mathbb{E}^2(f(G))$. For $\alpha \in (0,1]$, let

$$\mathcal{H}_\alpha = \Big\{ g : \mathbb{R}^d \to \mathbb{R} \text{ s.t. } \exists \beta \in [0,2), \lambda > 0, \forall x \in \mathbb{R}^d,\ |g(x)| \leq \lambda e^{|x|^\beta} \\ \forall x, y \in \mathbb{R}^d,\ |g(x) - g(y)| \leq \lambda e^{|x|^\beta \vee |y|^\beta} |x-y|^\alpha \Big\}.$$

Note that the assumptions of Theorem 3.2 are satisfied for $f \in \mathcal{H}_\alpha$ such that (1.1) holds and that the Hölder condition in the definition of $\mathcal{H}_\alpha$ implies the growth assumption for possibly larger constants λ and β.

Theorem 3.3 *Assume* (1.1), (1.3) *and that f admits a decomposition $f = f_1 + f_2 + \mathbf{1}_{\{d'=1\}} f_3$ with f_1 a C^1 function such that*

$$\forall M > 0,\ \mathbb{E}\left(\sup_{|\theta| \leq M} |f_1(G+\theta)| + \sup_{|\theta| \leq M} |\nabla f_1(G+\theta)| \right) < +\infty,$$

$f_2 \in \mathcal{H}_\alpha$ with $\alpha \in \left(\frac{\sqrt{d'^2+8d'} - d'}{4}, 1 \right]$ and $f_3 \in \mathcal{V}_A$. Then,

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

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 $d' \to \infty$. Theorem 3.3 follows from Propositions 3.7 and 3.14 below. With Proposition 2.2, one obtains the following corollary which enables us to construct confidence intervals for $\mathbb{E}(f(G))$ with our algorithm.

Corollary 3.4 *Under the assumptions of Theorem 3.3, if* $\mathrm{Var}(f(G)) > 0$*, then*

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

Remark 3.5 *When* $\mathrm{Var}(f(G))$ *is positive, then the optimal variance* $v^{f,A}(\vartheta_\star^{f,A}) - \mathbb{E}^2(f(G))$ *is also positive. The estimator* $v_n^{f,A}(\vartheta_n^{f,A}) - M_n^2(\theta_n^{f,A}, f)$ *converges a.s. to this variance but may take negative values for n small.*

Examples : Let us give some examples, inspired by financial applications, of functions f such that the hypotheses of Theorems 3.2 and 3.3 are satisfied.

- $f(x) = \left(K + \sum_{k=1}^d \omega_k e^{\sigma_k (Mx)_k}\right)^+$ where the coefficients K, ω_k and σ_k are real numbers and $M \in \mathbb{R}^{d\times d}$: this class of functions belonging to $\mathcal{H}_1$ includes the payoffs of Call and Put options written on baskets of underlyings in a multidimensional Black-Scholes framework or on a discretely sampled arithmetic average of a single Black-Scholes asset and the payoffs of exchange options on baskets.
- $f(x) = \left(K + \max_{k=1}^d \omega_k e^{\sigma_k (Mx)_k}\right)^+$, $f(x) = \left(K + \min_{k=1}^d \omega_k e^{\sigma_k (Mx)_k}\right)^+$: this class of functions belonging to $\mathcal{H}_1$ includes the payoffs of best-of options.
- when $d = 1$, the functions of bounded variation $f(x) = \mathbf{1}_{\{\omega e^{\sigma x} \geq K\}}$ and $f(x) = \mathbf{1}_{\{\omega e^{\sigma x} \leq K\}}$ belong to $\mathcal{V}_1$ and correspond respectively to binary Call and Put options in the Black-Scholes model.
- The time-discretization of the one-dimensional Black-Scholes model on the grid $0 = t_0 \leq t_1 \leq \ldots \leq t_d$ is given by $\varphi(G)$ where $\varphi(x) = (e^{(r-\frac{\sigma^2}{2})t_k + \sigma \sum_{j=1}^k x_j \sqrt{t_j - t_{j-1}}})_{1\leq k\leq d}$ with $\sigma > 0$. For the choice $A = (\sqrt{t_1}, \sqrt{t_2 - t_1}, \ldots, \sqrt{t_d - t_{d-1}})^*$ which corresponds to the Cameron-Martin formula for the underlying Brownian motion, each coordinate of the function φ is A-nondecreasing. Therefore, when $g : \mathbb{R}^d \to \mathbb{R}$ is either nondecreasing in each variable or nonincreasing in each variable, the function $g \circ \varphi$ is A-monotonic. For $g_1(y) = (y_d - K)^+$ and $g_2(y) = (y_d - K)^+ \mathbf{1}_{\{\min_k y_k \geq L\}}$, the functions $g_2 \circ \varphi$ and $(g_1 - g_2) \circ \varphi$, which correspond to the down-and-out and the down-and-in barrier Call options also belong to $\mathcal{V}_A$. More generally, all the barrier Call and Put option payoffs belong to $\mathcal{V}_A$.
- Let us consider the model

  $$dS_t = S_t\left(\sigma(t, S_t)dW_t + r dt\right),\ S_0 = s$$

  where $(W_t)_{t\geq 0}$ is a one-dimensional Brownian motion and the local volatility function $\sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ is bounded and such that $x \mapsto x\sigma(t,x)$ is Lipschitz continuous uniformly for $t \in [0,T]$. When discretizing this SDE by the Euler scheme with d steps of length $h = T/d$ on $[0,T]$, one approximates S_T by $\varphi(G)$ where

  $\varphi(x) = \phi_d(x_d, \phi_{d-1}(x_{d-1}, \ldots, \phi_1(x_1, s)))$ with $\phi_k(u,v) = v(1 + \sigma((k-1)h, v)\sqrt{h}u + rh)$, and $G = \dfrac{1}{\sqrt{h}}\left(W_h, W_{2h} - W_h, \ldots, W_{dh} - W_{(d-1)h}\right)$.

There exists $C > 0$ such that for all $k \in \{1, \dots, d\}$,

$$\forall u, v, u', v' \in \mathbb{R},\ |\phi_k(u,v)| \le C|v|(1+|u|)$$
$$|\phi_k(u,v) - \phi_k(u',v')| \le C\left((1 + (|u| \vee |u'|))|v - v'| + (|v| \vee |v'|)|u - u'|\right).$$

One deduces by induction that for $x, y \in \mathbb{R}^d$, $|\varphi(x)| \le C^d|s| \prod_{k=1}^d (1+|x_k|) \le C^d|s|e^{\sqrt{d}|x|}$ and

$$|\varphi(x) - \varphi(y)| \le C^d|s| \sum_{k=1}^d |x_k - y_k| \prod_{\substack{j=1 \\ j\neq k}}^d (1 + (|x_j| \vee |y_j|) \le C^d|s|\sqrt{d}e^{d(|x|\vee|y|)}|x-y|.$$

Hence, the functions $f(x) = (\varphi(x) - K)^+$ and $f(x) = (K - \varphi(x))^+$ corresponding to the Call and Put payoffs in the discretized model belong to $\mathcal{H}_1$.

We are now going to study the convergence properties of $M_n(\theta_n^{f,A}, g)$ in the multidimensional framework $d' \ge 1$ before obtaining stronger results in the case $d' = 1$ of a one-dimensional importance sampling parameter.

3.1 The general case

Proposition 3.6 *Let $(\theta_n)_{n\ge1}$ be a sequence of d-dimensional random vectors converging almost surely to some random vector θ_∞ and $g : \mathbb{R}^d \to \mathbb{R}$ be a continuous function such that $\forall M > 0$, $\mathbb{E}\left(\sup_{|\theta|\le M} |g(G+\theta)|\right) < \infty$. Then $M_n(\theta_n, g)$ converges a.s. to $\mathbb{E}(g(G))$.*

Proof : We apply Lemma 2.4 with $(X_i)_{i\ge1} = (G_i)_{i\ge1}$ and $h(\theta, x) = g(x+\theta)e^{-\theta.x - \frac{|\theta|^2}{2}}$. The continuity assumption follows from the continuity of g. Concerning the integrability condition, we deduce from (1.4) and the following inequality

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

that

$$\mathbb{E}\left(\sup_{|\theta|\le M}\left(|g(G+\theta)|e^{-\theta\cdot G - \frac{|\theta|^2}{2}}\right)\right) \le e^{\frac{dM^2}{2}} \sum_{\mu\in\{-M,M\}^d} \mathbb{E}\left(\sup_{|\theta|\le M}|g(G+\theta+\mu)|\right)$$
$$\le 2^d e^{\frac{dM^2}{2}} \mathbb{E}\left(\sup_{|\theta|\le(1+\sqrt{d})M}|g(G+\theta)|\right).$$

Therefore, a.s., $\theta \mapsto M_n(g,\theta)$ converges locally uniformly to the constant function $\theta \mapsto \mathbb{E}(h(\theta, G)) = \mathbb{E}(g(G))$. We easily conclude with the a.s. convergence of θ_n to θ_∞. ■

Proposition 3.7 *Assume that $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ is such that $\mathbb{E}(g^2(G+\theta_\star^{f,A})e^{-2\theta_\star^{f,A}.G}) < +\infty$ and admits a decomposition $g = g_1 + g_2$ with g_1 of class C^1 satisfying*

$$\forall M > 0,\ \mathbb{E}\left(\sup_{|\theta|\leq M} |g_1(\theta+G)| + \sup_{|\theta|\leq M} |\nabla g_1(\theta+G)|\right) < \infty \tag{3.2}$$

and $g_2 \in \mathcal{H}_\alpha$ for $\alpha \in \left(\frac{\sqrt{d'^2+8d'}-d'}{4}, 1\right]$. Then, under (1.1) and (1.3),

$$\sqrt{n}(M_n(\theta_n^{f,A}, g) - \mathbb{E}(g(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, \mathrm{Var}\left(g(G+\theta_\star^{f,A})e^{-\theta_\star^{f,A}.G-\frac{|\theta_\star^{f,A}|^2}{2}}\right)\right).$$

By the central limit theorem, $\sqrt{n}(M_n(\theta_\star^{f,A}, g) - \mathbb{E}(g(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, \mathrm{Var}\left(g(G+\theta_\star^{f,A})e^{-\theta_\star^{f,A}.G-\frac{|\theta_\star^{f,A}|^2}{2}}\right)\right)$. As a consequence, it is enough to check that for $i \in \{1,2\}$, $\sqrt{n}(M_n(\theta_n^{f,A}, g_i) - M_n(\theta_\star^{f,A}, g_i)) \xrightarrow{\mathrm{Pr}} 0$. The next lemma deals with the case $i = 1$.

Lemma 3.8 *Let $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a C^1 function satisfying (3.2). Then, under (1.1) and (1.3), $\sqrt{n}(M_n(\theta_n^{f,A}, g) - M_n(\theta_\star^{f,A}, g)) \xrightarrow{\mathrm{Pr}} 0$.*

Since, for $\varepsilon > 0$,

$$\mathbb{P}\left(\sqrt{n}|M_n(\theta_n^{f,A}, g_2) - M_n(\theta_\star^{f,A}, g_2)| \geq \varepsilon\right) \leq \mathbb{P}\left(n^\delta |\vartheta_n^{f,A} - \vartheta_\star^{f,A}| \geq 1\right) + \mathbb{P}\left(\sup_{|\vartheta-\vartheta_\star^{f,A}|\leq \frac{1}{n^\delta}} \sqrt{n}|M_n(A\vartheta, g_2) - M_n(A\vartheta_\star^{f,A}, g_2)| \geq \varepsilon\right),$$

choosing $\delta \in (d'/2\alpha(d'+2\alpha), 1/2)$, which is possible since $\alpha > \frac{\sqrt{d'^2+8d'}-d'}{4}$, the case $i = 2$ follows from the central limit theorem governing the convergence of $\vartheta_n^{f,A}$ to $\vartheta_\star^{f,A}$ (see Proposition 2.2) combined with the following result.

Proposition 3.9 *Let $\mathcal{A} \in \mathbb{R}^{d\times d'}$ and $g \in \mathcal{H}_\alpha$ for $\alpha \in (0,1]$,*

$$\forall \delta > \frac{d'}{2\alpha(d'+2\alpha)}, \forall \vartheta_0 \in \mathbb{R}^{d'}, \sup_{|\vartheta-\vartheta_0|\leq \frac{1}{n^\delta}} \sqrt{n}|M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| \xrightarrow{\mathrm{Pr}} 0.$$

Remark 3.10 *By the argument given in the case $i = 2$, if $g \in \mathcal{H}_\alpha$ for $\alpha > \frac{\sqrt{d'^2+8d'}-d'}{4}$,*

$$\sqrt{n}(M_n(\theta_{\nu_n}^{f,A}, g) - \mathbb{E}(g(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, \mathrm{Var}\left(g(G+\theta_\star^{f,A})e^{-\theta_\star^{f,A}.G-\frac{|\theta_\star^{f,A}|^2}{2}}\right)\right)$$

for any deterministic integer-valued sequence $(\nu_n)_n$ such that $\exists \lambda > 0,\ \exists \gamma > \frac{d'}{\alpha(d'+2\alpha)},\ \forall n \in \mathbb{N}^*,\ \nu_n \geq \lambda n^\gamma$.

Proof of Lemma 3.8 : The function $\theta \mapsto M_n(\cdot, g)$ is of class C^1 and it is easy to check that $\nabla_\theta M_n(\theta, g) = M_n(\theta, \bar{g})$ with $\bar{g}(x) = \nabla g(x) - g(x)x$. The mean value theorem gives $\sqrt{n}(M_n(\theta_n^{f,A}, g) - M_n(\theta_\star^{f,A}, g)) = A\sqrt{n}(\vartheta_n^{f,A} - \vartheta_\star^{f,A}).M_n(\bar{\theta}_n^{f,A}, \bar{g})$, with $\bar{\theta}_n^{f,A} \in (\theta_n^{f,A}, \theta_\star^{f,A})$. Since by Proposition 2.2 $\sqrt{n}(\vartheta_n^{f,A} - \vartheta_\star^{f,A})$ converges in law to a normal random variable, it is enough to prove that $M_n(\bar{\theta}_n^{f,A}, \bar{g}) \xrightarrow{\Pr} 0$. The a.s. convergence of $\vartheta_n^{f,A}$ to $\vartheta_\star^{f,A}$ implies the a.s. convergence of $\bar{\theta}_n^{f,A}$ to $\theta_\star^{f,A}$. Since

$$\sup_{|\theta|\leq M} |(G+\theta)g(G+\theta)| \leq \left(\sum_{k=1}^{d}(e^{G^k} + e^{-G^k}) + M\right) \sup_{|\theta|\leq M} |g(G+\theta)|,$$

(3.2) combined with the reasoning made at the beginning of the proof of Proposition 3.6 yields

$$\forall M > 0,\ \mathbb{E}\left(\sup_{|\theta|\leq M} |(G+\theta)g(G+\theta)| + \sup_{|\theta|\leq M} |\nabla g(G+\theta)|\right) < +\infty. \tag{3.3}$$

Then, Proposition 3.6 implies that $M_n(\bar{\theta}_n^{f,A}, \bar{g}) \xrightarrow{\Pr} \mathbb{E}(\bar{g}(G))$. By (3.3) and the reasoning made at the beginning of the proof of Proposition 3.6,

$$\forall M > 0, \mathbb{E}\left(\sup_{|\theta|\leq M} |\bar{g}(G+\theta)|e^{-\theta\cdot G - \frac{|\theta|^2}{2}}\right) < +\infty.$$

Hence, Lebesgue's Theorem implies that $\nabla_\theta \mathbb{E}\left(g(G+\theta)e^{-\theta\cdot G-\frac{|\theta|^2}{2}}\right) = \mathbb{E}\left(\bar{g}(G+\theta)e^{-\theta\cdot G-\frac{|\theta|^2}{2}}\right)$. Since the left-hand-side is equal to 0, one deduces for $\theta = 0$ that $\mathbb{E}(\bar{g}(G)) = 0$. ■

Remark 3.11 *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a C^2 function, $\bar{g}(x) = \nabla g(x) - g(x)x$ and $\bar{\bar{g}}(x) \stackrel{\text{def}}{=} \nabla^2 g(x) - g(x)I_d - x\nabla^* g(x) - \nabla g(x)x^* + g(x)xx^*$. Assume that*

$$\mathbb{E}\left((|g(\theta_\star^{f,A} + G)|^2 + |\bar{g}(\theta_\star^{f,A} + G)|^2)e^{-2\theta_\star^{f,A}\cdot G}\right) < +\infty \tag{3.4}$$

$$\forall M > 0, \mathbb{E}\left(\sup_{|\theta|\leq M} |g(\theta+G)| + \sup_{|\theta|\leq M} |\nabla g(\theta+G)| + \sup_{|\theta|\leq M} |\nabla^2 g(\theta+G)|\right) < \infty. \tag{3.5}$$

Let $(\nu_n)_n$ be a deterministic integer valued sequence such that $\exists \lambda > 0,\ \forall n \in \mathbb{N}^*,\ \nu_n \geq \lambda\sqrt{n}$. Then, using the decomposition

$$\begin{aligned}\sqrt{n}(M_n(\theta_{\nu_n}^{f,A}, g) - M_n(\theta_\star^{f,A}, g)) &= \frac{1}{\sqrt{\nu_n}}\sqrt{n}M_n(\theta_\star^{f,A}, \bar{g}) \cdot \sqrt{\nu_n}(\theta_{\nu_n}^{f,A} - \theta_\star^{f,A}) \\ &+ \frac{\sqrt{n}}{\nu_n}\sqrt{\nu_n}(\theta_{\nu_n}^{f,A} - \theta_\star^{f,A})^* \left(\int_0^1 (1-t)M_n(t\theta_{\nu_n}^{f,A} + (1-t)\theta_\star^{f,A}, \bar{\bar{g}})dt\right)\sqrt{\nu_n}(\theta_{\nu_n}^{f,A} - \theta_\star^{f,A}),\end{aligned}$$

one obtains that under (1.1) *and* (1.3)*, the left-hand-side converges in probability to* 0*. As a consequence, $\sqrt{n}(M_n(\theta_{\nu_n}^{f,A}, g) - \mathbb{E}(g(G))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, \mathrm{Var}\left(g(G+\theta_\star^{f,A})e^{-\theta_\star^{f,A}\cdot G - \frac{|\theta_\star^{f,A}|^2}{2}}\right)\right)$. More generally, if g is of class C^k and satisfies moment assumptions like* (3.4) *and* (3.5) *respectively involving its derivatives up to order $k-1$ and k, this result is preserved if $\exists \lambda > 0,\ \forall n \in \mathbb{N}^*,\ \nu_n \geq \lambda n^{1/k}$.*

In order to prove Proposition 3.9, we need the following Lemma :

Lemma 3.12 *If $g \in \mathcal{H}_\alpha$ for $\alpha \in (0,1]$, then*

$$\forall M>0,\ \exists C>0,\ \forall \theta,\theta' \in \bar{B}(0,M),\ \forall n \in \mathbb{N}^*,\ \mathbb{E}\left(\left(M_n(\theta,g)-M_n(\theta',g)\right)^2\right) \leq \frac{C|\theta-\theta'|^{2\alpha}}{n}.$$

Proof : Let $M>0$. Since, by (1.4),

$$\mathbb{E}\left(g(G+\theta)e^{-\theta G-\frac{\theta^2}{2}}-g(G+\theta')e^{-\theta' G-\frac{\theta'^2}{2}}\right)=0,$$

it is enough to check that

$$\exists C>0,\ \forall \theta,\theta' \in \bar{B}(0,M),\ \mathbb{E}\left(\left(g(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}-g(G+\theta')e^{-\theta'.G-\frac{|\theta'|^2}{2}}\right)^2\right) \leq C|\theta-\theta'|^{2\alpha}.$$

One has

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

Let $\lambda>0$ and $\beta \in [0,2)$ be such that

$$\forall x \in \mathbb{R}^d,\ |g(x)| \leq \lambda e^{|x|^\beta} \tag{3.6}$$

$$\forall x,y \in \mathbb{R}^d,\ |g(x)-g(y)| \leq \lambda e^{|x|^\beta \vee |y|^\beta}|x-y|^\alpha. \tag{3.7}$$

One has

$$\text{for } c=2^{(\beta-1)^+},\ \forall a,b \geq 0,\ (a+b)^\beta \leq c(a^\beta+b^\beta). \tag{3.8}$$

Since for $\theta \in \bar{B}(0,M)$, $|\nabla_\theta e^{-\theta G-\frac{\theta^2}{2}}| = |(G+\theta)e^{-\theta G-\frac{\theta^2}{2}}| \leq (|G|+M)e^{M|G|}$, one deduces that for $\theta,\theta' \in \bar{B}(0,M)$,

$$\begin{aligned}\mathbb{E}\left(\left(g(G+\theta)e^{-\theta.G-\frac{|\theta|^2}{2}}-g(G+\theta')e^{-\theta'.G-\frac{|\theta'|^2}{2}}\right)^2\right) &\leq 2\lambda^2 e^{2cM^\beta}\mathbb{E}\left(\left(|\theta-\theta'|^{2\alpha}+|\theta-\theta'|^2(|G|+M)^2\right)e^{2M|G|+2c|G|^\beta}\right)\\ &\leq C|\theta-\theta'|^{2\alpha}.\end{aligned}$$

■

Proof of Proposition 3.9 : Let $\varepsilon>0$.

$$\mathbb{P}\left(\sup_{|\vartheta-\vartheta_0|\leq\frac{1}{n^\delta}}\sqrt{n}|M_n(\mathcal{A}\vartheta,g)-M_n(\mathcal{A}\vartheta_0,g)|>\varepsilon\right) \leq n\mathbb{P}\left(|G|>\sqrt{2d\log n}\right) + \mathbb{P}\left(\sup_{|\vartheta-\vartheta_0|\leq\frac{1}{n^\delta}}\sqrt{n}|M_n(\mathcal{A}\vartheta,g)-M_n(\mathcal{A}\vartheta_0,g)|>\varepsilon,\ \max_{1\leq i\leq n}|G_i| \leq \sqrt{2d\log n}\right).$$

Since

$$\mathbb{P}(|G| > \sqrt{2d\log n}) \leq \sum_{k=1}^{d} \mathbb{P}\left(|G^k| > \sqrt{2\log n}\right)$$
$$= 2d\mathbb{P}\left(G^1 > \sqrt{2\log n}\right) \leq \frac{2d}{\sqrt{2\log n}} e^{-\frac{(\sqrt{2\log n})^2}{2}}, \tag{3.9}$$

the second term of the r.h.s tends to 0 as n goes to infinity. Now, let us focus on the first term.

Let $M = |\vartheta_0| + 1$ and $\tilde{M} = |\mathcal{A}|M$. For $\vartheta', \vartheta \in \bar{B}(0, M)$, using (3.7), (3.8) and (3.6) for the second inequality, one obtains

$$|M_n(\mathcal{A}\vartheta', g) - M_n(\mathcal{A}\vartheta, g)| \leq \frac{1}{n}\sum_{i=1}^{n} |g(G_i + \mathcal{A}\vartheta') - g(G_i + \mathcal{A}\vartheta)| e^{-\mathcal{A}\vartheta'\cdot G_i - \frac{|\mathcal{A}\vartheta'|^2}{2}}$$
$$+ \frac{1}{n}\sum_{i=1}^{n} |g(G_i + \mathcal{A}\vartheta)| \left| e^{-\mathcal{A}\vartheta'\cdot G_i - \frac{|\mathcal{A}\vartheta'|^2}{2}} - e^{-\mathcal{A}\vartheta\cdot G_i - \frac{|\mathcal{A}\vartheta|^2}{2}} \right|$$
$$\leq \frac{\lambda|\mathcal{A}|^\alpha|\vartheta' - \vartheta|^\alpha}{n} \sum_{i=1}^{n} \mathrm{e}^{c(|G_i|^\beta + \tilde{M}^\beta) + \tilde{M}|G_i|} \left(1 + (2\tilde{M})^{1-\alpha}(\tilde{M} + |G_i|)\right).$$

Hence, when $\max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n}$ there exists a constant γ not depending on n such that if $\nu \stackrel{\text{def}}{=} \frac{\beta\vee 1}{2}$,

$$\forall \vartheta', \vartheta \in \bar{B}(0, M), \quad |M_n(\mathcal{A}\vartheta', g) - M_n(\mathcal{A}\vartheta, g)| \leq \gamma|\vartheta - \vartheta'|^\alpha e^{\gamma(\log n)^\nu}. \tag{3.10}$$

We can cover $\bar{B}(\vartheta_0, \frac{1}{n^\delta})$ with $K = \mathcal{C}\lceil(\gamma^{\frac{1}{\alpha}} n^{\frac{1}{2\alpha}-\delta} e^{\frac{\gamma}{\alpha}(\log n)^\nu}/\varepsilon^{\frac{1}{\alpha}})^{d'}\rceil$ balls of radius $\left(\frac{\varepsilon}{2\gamma e^{\gamma(\log n)^\nu}\sqrt{n}}\right)^{\frac{1}{\alpha}}$, where $\mathcal{C}$ is a geometrical constant not depending on n. For $k \in \{1, \ldots, K\}$, let B_k denote the $k-th$ ball and ϑ_k its center. By (3.10), when $\max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n}$,

$$\forall k \in \{1, \ldots, K\}, \ \sup_{\vartheta\in B_k} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_k, g)| \leq \frac{\varepsilon}{2\sqrt{n}}.$$

Using Lemma 3.12 for the fourth inequality, one deduces that

$$\mathbb{P}\left(\sup_{|\vartheta-\vartheta_0|\leq\frac{1}{n^\delta}} \sqrt{n}\,|M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| > \varepsilon, \ \max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n}\right)$$
$$\leq \mathbb{P}\Bigg(\exists k \leq K : |M_n(\mathcal{A}\vartheta_k, g) - M_n(\mathcal{A}\vartheta_0, g)| > \frac{\varepsilon}{\sqrt{n}} - \sup_{\vartheta\in B_k} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_k, g)|$$
$$, \ \max_{1\leq i\leq n} |G_i| \leq \sqrt{2d\log n}\Bigg)$$
$$\leq \mathbb{P}\left(\max_{k\leq K} |M_n(\mathcal{A}\vartheta_k, g) - M_n(\mathcal{A}\vartheta_0, g)| > \frac{\varepsilon}{2\sqrt{n}}\right) \leq \sum_{k\leq K} \frac{4n}{\varepsilon^2}\mathbb{E}\left((M_n(\mathcal{A}\vartheta_k, g) - M_n(\mathcal{A}\vartheta_0, g))^2\right)$$
$$\leq \sum_{k\leq K} \frac{4n}{\varepsilon^2}\frac{C|\vartheta - \vartheta_k|^{2\alpha}}{n} \leq Cn^{\frac{d'}{2\alpha}-(d'+2\alpha)\delta} e^{\frac{d'\gamma}{\alpha}(\log n)^\nu}. \tag{3.11}$$

Since $\beta < 2$ and $\delta > \frac{d'}{2\alpha(d'+2\alpha)}$, $\nu < 1$ and $\frac{d'}{2\alpha} - (d'+2\alpha)\delta < 0$. Therefore, the upper bound in Equation (3.11) converges to 0 as n increases to infinity. ■

3.2 The case of a one-dimensional importance sampling parameter

In the present section, dedicated to the case $d' = 1$ of a one-dimensional importance sampling parameter, we obtain convergence results under weaker assumptions on the function g.

Proposition 3.13 *Let $\mathcal{A} \in \mathbb{R}^d$. Assume that $g : \mathbb{R}^d \to \mathbb{R}$ admits a decomposition $g = g_1 + g_2$ with g_1 a continuous function such that $\forall M > 0$, $\mathbb{E}\left(\sup_{|\theta|\leq M} |g_1(G+\theta)|\right) < +\infty$ and $g_2 \in \mathcal{V}_{\mathcal{A}}$. Then, for any sequence $(\vartheta_n)_n$ of real valued random variables converging a.s. to some deterministic limit $\vartheta_\star \in \mathbb{R}$, $M_n(\mathcal{A}\vartheta_n, g)$ converges a.s. to $\mathbb{E}(g(G))$.*

Proof of Proposition 3.13 : By Proposition 3.6, it is enough to deal with the situation where $g = g_\uparrow + g_\downarrow$ with $g_\uparrow$ (resp. $g_\downarrow$) an $\mathcal{A}$-nondecreasing (resp. $\mathcal{A}$-nonincreasing) function satisfying (3.1). One has $g = g_\uparrow \mathbf{1}_{\{g_\uparrow \geq 0\}} + g_\uparrow \mathbf{1}_{\{g_\uparrow < 0\}} + g_\downarrow \mathbf{1}_{\{g_\downarrow \geq 0\}} + g_\downarrow \mathbf{1}_{\{g_\downarrow < 0\}}$ where the functions $g_\uparrow \mathbf{1}_{\{g_\uparrow \geq 0\}}$ and $-g_\downarrow \mathbf{1}_{\{g_\downarrow < 0\}}$ (resp. $g_\downarrow \mathbf{1}_{\{g_\downarrow \geq 0\}}$ and $-g_\uparrow \mathbf{1}_{\{g_\uparrow < 0\}}$) are nonnegative, $\mathcal{A}$-nondecreasing (resp. $\mathcal{A}$-nonincreasing) and satisfy (3.1). As a consequence, it is enough to deal with the case g nonnegative, $\mathcal{A}$-monotonic and satisfying (3.1). Choosing $\vartheta' \geq \vartheta$ when g is $\mathcal{A}$-nondecreasing and $\vartheta \geq \vartheta'$ when g is $\mathcal{A}$-nonincreasing, one has for all $x \in \mathbb{R}^d$,

$$
\begin{aligned}
&g(x+\mathcal{A}\vartheta')e^{-\mathcal{A}\vartheta'\cdot x - \frac{|\mathcal{A}\vartheta'|^2}{2}} - g(x+\mathcal{A}\vartheta)e^{-\mathcal{A}\vartheta\cdot x - \frac{|\mathcal{A}\vartheta|^2}{2}} \\
&\geq \left(g(x+\mathcal{A}\vartheta')(e^{-\mathcal{A}\vartheta'\cdot x - \frac{|\mathcal{A}\vartheta'|^2}{2}} - e^{-\mathcal{A}\vartheta\cdot x - \frac{|\mathcal{A}\vartheta|^2}{2}})\right) \vee \left(g(x+\mathcal{A}\vartheta)(e^{-\mathcal{A}\vartheta'\cdot x - \frac{|\mathcal{A}\vartheta'|^2}{2}} - e^{-\mathcal{A}\vartheta\cdot x - \frac{|\mathcal{A}\vartheta|^2}{2}})\right).
\end{aligned} \tag{3.12}
$$

From now on, we suppose that g is nonnegative, $\mathcal{A}$-nondecreasing and satisfies (3.1) : a symmetric argument applies to the nonincreasing case. Let $\varepsilon > 0$, $\eta \in (0,1]$. For $m \in \mathbb{N}^*$,

$$
\begin{aligned}
\mathbb{P}\Big(\exists n \geq m,\, |\mathbb{E}(g(G)) - M_n(\mathcal{A}\vartheta_n, g)| \geq \varepsilon\Big) &\leq \mathbb{P}\left(\exists n \geq m,\ |\mathbb{E}(g(G)) - M_n(\mathcal{A}\vartheta_\star, g)| \geq \frac{\varepsilon}{2}\right) \\
&\quad + \mathbb{P}\left(\exists n \geq m,\ |\vartheta_n - \vartheta_\star| > \eta\right) \\
&\quad + \mathbb{P}\left(\forall n \geq m,\ |\vartheta_n - \vartheta_\star| \leq \eta \quad \text{and} \quad \exists n \geq m,\ |M_n(\mathcal{A}\vartheta_\star, g) - M_n(\mathcal{A}\vartheta_n, g)| \geq \frac{\varepsilon}{2}\right).
\end{aligned}
$$

By the strong law of large numbers and the a.s. convergence of ϑ_n to $\vartheta_\star$, the first two terms on the right-hand-side both converge to 0 as $m \to +\infty$. Let us check that the third one also converges to 0. Let $M = |\vartheta_\star| + 1$, $K > 0$. For $-M \leq \vartheta \leq \vartheta' \leq M$, one has using (3.12) for

the first inequality then (3.1) and (3.8),

$$
\begin{aligned}
&M_n(\mathcal{A}\vartheta',g)-M_n(\mathcal{A}\vartheta,g)\\
&\qquad\geq -\frac{1}{n}\sum_{i=1}^{n}(|g(G_i+\mathcal{A}\vartheta')|\wedge|g(G_i+\mathcal{A}\vartheta)|)\left|e^{-\mathcal{A}\vartheta'\cdot G_i-\frac{|\mathcal{A}\vartheta'|^2}{2}}-e^{-\mathcal{A}\vartheta\cdot G_i-\frac{|\mathcal{A}\vartheta|^2}{2}}\right|\mathbf{1}_{\{|G_i|\leq K\}}\\
&\qquad\quad-\frac{1}{n}\sum_{i=1}^{n}\left(|g(G_i+\mathcal{A}\vartheta')|e^{-\mathcal{A}\vartheta'\cdot G_i-\frac{|\mathcal{A}\vartheta'|^2}{2}}+|g(G_i+\mathcal{A}\vartheta)|e^{-\mathcal{A}\vartheta\cdot G_i-\frac{|\mathcal{A}\vartheta|^2}{2}}\right)\mathbf{1}_{\{|G_i|>K\}}\\
&\qquad\geq -\gamma_K(\vartheta'-\vartheta)-\frac{C}{n}\sum_{i=1}^{n}e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}},
\end{aligned}
$$

where $\gamma_K=\lambda|\mathcal{A}|e^{c(K^\beta+(M|\mathcal{A}|)^\beta)}(M|\mathcal{A}|+K)e^{MK|\mathcal{A}|}$ and $C=2\lambda e^{c(M|\mathcal{A}|)^\beta}$. When $|\vartheta_n-\vartheta_\star|\leq\eta$, choosing $\vartheta=\vartheta_n$ and $\vartheta'=\vartheta_\star+\eta$ then $\vartheta=\vartheta_\star-\eta$ and $\vartheta'=\vartheta_n$, one deduces that $M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}\vartheta_n,g)$ is bounded from below and from above respectively by

$$
M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star+\eta),g)-\gamma_K(\vartheta_\star+\eta-\vartheta_n)-\frac{C}{n}\sum_{i=1}^{n}e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}}
$$

$$
\text{and } M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star-\eta),g)+\gamma_K(\vartheta_n+\eta-\vartheta_\star)+\frac{C}{n}\sum_{i=1}^{n}e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}}.
$$

Choosing K such that $\mathbb{E}\left(e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}}\right)\leq\frac{\varepsilon}{8C}$ and then η such that $2\gamma_K\eta\leq\frac{\varepsilon}{8}$, we deduce that

$$
\begin{aligned}
&\mathbb{P}\Big(\forall n\geq m,\ |\vartheta_n-\vartheta_\star|\leq\eta\quad\text{and}\quad\exists n\geq m,\ |M_n(\mathcal{A}\vartheta_\star,h)-M_n(\mathcal{A}\vartheta_n,h)|\geq\frac{\varepsilon}{2}\Big)\\
&\qquad\leq\mathbb{P}\left(\exists n\geq m,\ \frac{1}{n}\sum_{i=1}^{n}e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}}\geq\frac{\varepsilon}{4C}\right)\\
&\qquad\quad+\mathbb{P}\left(\exists n\geq m,\ M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star+\eta),g)\leq-\frac{\varepsilon}{8}\right)\\
&\qquad\quad+\mathbb{P}\left(\exists n\geq m,\ M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star-\eta),g)\geq\frac{\varepsilon}{8}\right).
\end{aligned}
$$

By the strong law of large numbers $M_n(\mathcal{A}\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star+\eta),g)$ and $M_n(\vartheta_\star,g)-M_n(\mathcal{A}(\vartheta_\star-\eta),g)$ both converge a.s. to 0 and $\frac{1}{n}\sum_{i=1}^{n}e^{c|G_i|^\beta+M|\mathcal{A}||G_i|}\mathbf{1}_{\{|G_i|>K\}}$ to some limit not greater than $\frac{\varepsilon}{8C}$. One concludes that each term on the right-hand-side converges to 0 as $m\to\infty$. ■

Proposition 3.14 *Assume that $g:\mathbb{R}^d\longrightarrow\mathbb{R}$ is such that $\mathbb{E}(g^2(G+\theta_\star^{f,A})e^{-2\theta_\star^{f,A}\cdot G})<+\infty$ and admits a decomposition $g=g_1+g_2+\mathbf{1}_{\{d'=1\}}g_3$ with g_1 of class C^1 satisfying (3.2), $g_2\in\mathcal{H}_\alpha$ for $\alpha\in\left(\frac{\sqrt{d'^2+8d'}-d'}{4},1\right]$ and $g_3\in\mathcal{V}_A$. Then, under (1.1) and (1.3),*

$$
\sqrt{n}(M_n(\theta_n^{f,A},g)-\mathbb{E}(g(G)))\xrightarrow{\mathcal{L}}\mathcal{N}_1\left(0,\mathrm{Var}\left(g(G+\theta_\star^{f,A})e^{-\theta_\star^{f,A}.G-\frac{|\theta_\star^{f,A}|^2}{2}}\right)\right).
$$

As in Proposition 3.7, this statement is proved by combining the usual central limit theorem governing the convergence of $\sqrt{n}(M_n(\theta_\star^{f,A},g)-\mathbb{E}(g(G)))$, Lemma 3.8, Proposition 3.9, the

decomposition of functions in $\mathcal{V}_A$ given at the beginning of the proof of Proposition 3.13 and the next result.

Proposition 3.15 *Let $\mathcal{A} \in \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}$ be an $\mathcal{A}$-monotonic function with constant sign satisfying* (3.1)*,*

$$\forall \delta > 1/4,\ \forall \vartheta_0 \in \mathbb{R},\quad \sup_{\vartheta \in [\vartheta_0 \pm \frac{1}{n^\delta}]} \sqrt{n} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| \stackrel{\Pr}{\to} 0.$$

Remark 3.16 *Assume that $d' = 1$. Let $g \in \mathcal{V}_A$, and $(\nu_n)_n$ be a deterministic integer valued sequence such that*

$$\exists \lambda > 0,\ \exists \gamma > \frac{1}{2},\ \forall n \in \mathbb{N}^*,\ \nu_n \geq \lambda n^\gamma.$$

Combining Propositions 2.2 and 3.15, one obtains that under (1.1) *and* (1.3)*,* $\sqrt{n}(M_n(\theta_{\nu_n}^{f,A}, g) - \mathbb{E}(g(G))) \stackrel{\mathcal{L}}{\to} \mathcal{N}_1\left(0, \mathrm{Var}\left(g(G + \theta_\star^{f,A}) e^{-\theta_\star^{f,A}.G - \frac{|\theta_\star^{f,A}|^2}{2}}\right)\right)$.

Proof : Up to a multiplication by -1, we may assume that g is nonnegative. Moreover, we only deal with the case g $\mathcal{A}$-nondecreasing, the nonincreasing case being obtained by a symmetric argument. By (3.12), for $\vartheta' < \vartheta''$ and $\vartheta \in [\vartheta', \vartheta'']$

$$M_n(\mathcal{A}\vartheta', g) - \frac{1}{n}\sum_{i=1}^n |g(G_i + \mathcal{A}\vartheta)| \left| e^{-\mathcal{A}\vartheta \cdot G_i - \frac{|\mathcal{A}\vartheta|^2}{2}} - e^{-\mathcal{A}\vartheta' \cdot G_i - \frac{|\mathcal{A}\vartheta'|^2}{2}} \right|$$
$$\leq M_n(\mathcal{A}\vartheta, g) \leq M_n(\mathcal{A}\vartheta'', h) + \frac{1}{n}\sum_{i=1}^n |g(G_i + \mathcal{A}\vartheta)| \left| e^{-\mathcal{A}\vartheta \cdot G_i - \frac{|\mathcal{A}\vartheta|^2}{2}} - e^{-\mathcal{A}\vartheta'' \cdot G_i - \frac{|\mathcal{A}\vartheta''|^2}{2}} \right|.$$

With (3.1) and (3.8), one deduces that if $-M \leq \vartheta' \leq \vartheta'' \leq M$,

$$\sup_{\vartheta \in [\vartheta', \vartheta'']} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| \leq \max(|M_n(\mathcal{A}\vartheta', g) - M_n(\mathcal{A}\vartheta_0, g)|, |M_n(\mathcal{A}\vartheta'', g) - M_n(\mathcal{A}\vartheta_0, g)|)$$
$$+ \frac{C(\vartheta'' - \vartheta')}{n} \sum_{i=1}^n e^{c|G_i|^\beta + M|\mathcal{A}||G_i|}(M|\mathcal{A}| + |G_i|). \quad (3.13)$$

Let $\nu = \frac{\beta \vee 1}{2}$ and $M = |\vartheta_0| + 1$. When $\max_{1 \leq i \leq n} |G_i| \leq \sqrt{2d \log n}$, the second term on the r.h.s. is smaller than $\gamma e^{\gamma (\log n)^\nu}(\vartheta'' - \vartheta')$ where the constant γ does not depend on n. Let $\varepsilon > 0$. We set $K = \lceil 2\gamma n^{\frac{1}{2} - \delta} e^{\gamma (\log n)^\nu} / \varepsilon \rceil$ and $\vartheta_k = \vartheta_0 + k\varepsilon / 2\gamma e^{\gamma (\log n)^\nu}$ for $k \in \{-K, \ldots, K\}$. Applying (3.13) with $\vartheta' = \vartheta_k$ and $\vartheta'' = \vartheta_{k+1}$, one obtains that when $\max_{1 \leq i \leq n} |G_i| \leq \sqrt{2d \log n}$,

$$\sup_{\vartheta \in [\vartheta_k, \vartheta_{k+1}]} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| \leq \frac{\varepsilon}{2\sqrt{n}}$$
$$+ \max(|M_n(\mathcal{A}\vartheta_k, g) - M_n(\mathcal{A}\vartheta_0, g)|, |M_n(\mathcal{A}\vartheta_{k+1}, g) - M_n(\mathcal{A}\vartheta_0, g)|).$$

Therefore,

$$\mathbb{P}\left(\sup_{\vartheta \in [\vartheta_0 \pm \frac{1}{n^\delta}]} |M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta_0, g)| \geq \frac{\varepsilon}{\sqrt{n}} \right) \leq \mathbb{P}\left(\max_{1 \leq i \leq n} |G_i| > \sqrt{2d \log n} \right)$$
$$+ \mathbb{P}\left(\max_{1 \leq i \leq n} |G_i| \leq \sqrt{2d \log n}, \max_{|k| \leq K} |M_n(\mathcal{A}\vartheta_k, g) - M_n(\mathcal{A}\vartheta_0, g)| \geq \frac{\varepsilon}{2\sqrt{n}} \right).$$

By (3.9), the first term on the right-hand-side tends to 0 as $n \to \infty$. Reasoning like in the end of the proof of Proposition 3.9, with the next lemma replacing Lemma 3.12, one concludes that the second term also tends to 0. ■

Lemma 3.17 *When $\mathcal{A} \in \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}$ is a $\mathcal{A}$-monotonic function with constant sign satisfying* (3.1)*,*

$$\forall M > 0, \exists C > 0,\ \forall \vartheta, \vartheta' \in [-M, M],\ \forall n \in \mathbb{N}^*,\ \mathbb{E}\left(\left(M_n(\mathcal{A}\vartheta, g) - M_n(\mathcal{A}\vartheta', g)\right)^2\right) \leq \frac{C|\vartheta - \vartheta'|}{n}.$$

Proof : Choosing $\vartheta' \geq \vartheta$ if g is nonnegative and $\mathcal{A}$-nondecreasing or nonpositive and $\mathcal{A}$-nonincreasing and $\vartheta \geq \vartheta'$ otherwise, one has

$$\begin{aligned}
&\mathbb{E}\left(\left(g(G + \mathcal{A}\vartheta)e^{-\mathcal{A}\vartheta.G - \frac{|\mathcal{A}\vartheta|^2}{2}} - g(G + \mathcal{A}\vartheta')e^{-\mathcal{A}\vartheta'.G - \frac{|\mathcal{A}\vartheta'|^2}{2}}\right)^2\right) \\
&\quad = \mathbb{E}\left(g^2(G)e^{-\mathcal{A}\vartheta.G + \frac{|\mathcal{A}\vartheta|^2}{2}}\right) + \mathbb{E}\left(g^2(G)e^{-\mathcal{A}\vartheta'.G + \frac{|\mathcal{A}\vartheta'|^2}{2}}\right) \\
&\qquad\qquad - 2\mathbb{E}\left(g(G)g(G + \mathcal{A}(\vartheta' - \vartheta))e^{-\mathcal{A}\vartheta'.G + \mathcal{A}\vartheta.\mathcal{A}\vartheta' - \frac{|\mathcal{A}\vartheta'|^2}{2}}\right) \\
&\quad \leq \mathbb{E}\left(g^2(G)\left(e^{-\mathcal{A}\vartheta.G + \frac{|\mathcal{A}\vartheta|^2}{2}} + e^{-\mathcal{A}\vartheta'.G + \frac{|\mathcal{A}\vartheta'|^2}{2}} - 2e^{-\mathcal{A}\vartheta'.G + \mathcal{A}\vartheta.\mathcal{A}\vartheta' - \frac{|\mathcal{A}\vartheta'|^2}{2}}\right)\right).
\end{aligned}$$

Then, the conclusion is a consequence of the following inequality : for $\theta, \theta' \in \mathbb{R}^d$ with $|\theta| \vee |\theta'| \leq |\mathcal{A}|M$,

$$\begin{aligned}
&\mathbb{E}\left(g^2(G)\left(e^{-\theta.G + \frac{|\theta|^2}{2}} + e^{-\theta'.G + \frac{|\theta'|^2}{2}} - 2e^{-\theta'.G + \theta.\theta' - \frac{|\theta'|^2}{2}}\right)\right) \\
&\quad \leq C\mathbb{E}\left(e^{\frac{|G|^2}{4}}\left(\left|e^{-\theta.G + \frac{|\theta|^2}{2}} - e^{-\theta'.G + \frac{|\theta'|^2}{2}}\right| + 2e^{-\theta'.G - \frac{|\theta'|^2}{2}}\left|e^{|\theta'|^2} - e^{\theta.\theta'}\right|\right)\right) \\
&\quad \leq C\left(|\theta - \theta'|\int_0^1 e^{\frac{3|\vartheta(t)|^2}{2}}\int_{\mathbb{R}^d}|\vartheta(t) - x|e^{-\frac{|x + 2\vartheta(t)|^2}{4}}dxdt + 2e^{\frac{|\theta'|^2}{2}}|e^{|\theta'|^2} - e^{\theta.\theta'}|\int_{\mathbb{R}^d}e^{-\frac{|x + 2\theta'|^2}{4}}dx\right) \\
&\quad \leq C|\theta - \theta'|.
\end{aligned}$$

■

4 Generalization

Let X be a $q-$dimensional random variable and $h : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}$ and $\phi : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}$ two measurable functions such that a.s. $\theta \mapsto \phi(\theta, X)$ is a convex (and therefore continuous by finiteness of ϕ) function bounded from below by some deterministic finite constant and $\mathbb{E}(\phi(\theta_0, X)) < \infty$ for some $\theta_0 \in \mathbb{R}^d$. We are interested in computing

$$\begin{cases} \mathbb{E}(h(\theta_\star^\phi, X)) \\ \text{with} \quad \mathbb{E}(\phi(\theta_\star^\phi, X)) = \inf_\theta \mathbb{E}(\phi(\theta, X)). \end{cases} \tag{4.1}$$

For any function $\psi : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}$, we define $m_n(\theta, \psi) = \frac{1}{n}\sum_{i=1}^n \psi(\theta, X_i)$, where $(X_i)_{i\geq 1}$ is an i.i.d. sample from the distribution of X. To solve the general Problem (4.1), we advise to compute instead the sample approximation

$$\begin{cases} m_n(\theta_n^\phi, h) \\ \text{with} \quad m_n(\theta_n^\phi, \phi) = \inf_{\theta\in\mathbb{R}^d} m_n(\theta, \phi) \end{cases}$$

The first part of this work was devoted to the particular case where $q = d$, $h(\theta, x) = f(\theta + x)\,\mathrm{e}^{-\theta\cdot x - \frac{|\theta|^2}{2}}$ and $\phi(\theta, x) = f^2(x)\,\mathrm{e}^{-\theta\cdot x + \frac{|\theta|^2}{2}}$ for a given function $f : \mathbb{R}^d \to \mathbb{R}^d$ and a standard normal random vector X. In the same context of importance sampling, when $q = d = 1$, and X is a gamma random variable with density $\mathbf{1}_{\{x>0\}}\frac{x^{a-1}e^{-x}}{\Gamma(a)}$ $(a > 0)$, for $f : \mathbb{R} \to \mathbb{R}$ with polynomial growth,

$$\forall\theta > -1,\ \mathbb{E}(f(X)) = \mathbb{E}\left((1+\theta)^a f((1+\theta)X)e^{-\theta X}\right).$$

Minimizing the variance of the random variable on the right-hand-side is equivalent to minimizing $\mathbb{E}\left(\phi(\theta, X)\right)$ where $\phi(\theta, x) \stackrel{\text{def}}{=} (1+\theta)^a f^2(x)e^{-\frac{\theta x}{1+\theta}}$. When $a \geq 1$, the function $\theta \mapsto \phi(\theta, X)$ is a.s. convex since $\frac{\partial^2\phi}{\partial\theta^2}(\theta, x) = (1+\theta)^{a-4}\left[(a-1)(1+\theta)^2 + ((a-1)(1+\theta)+x)^2\right]f^2(x)e^{-\frac{\theta x}{1+\theta}}$. The asymptotic properties of $m_n(\theta_n^\phi, h)$ where $h(\theta, x) = (1+\theta)^a f((1+\theta)x)e^{-\theta x}$ may be deduced from what follows.

By the assumptions made on ϕ, a.s. $\theta \to m_n(\theta, \phi)$ is continuous and, by Fatou's Lemma, $\theta \mapsto \mathbb{E}(\phi(\theta, X))$ is lower semi continuous. In order to ensure the existence and uniqueness of $\theta_\star^\phi$ and θ_n^ϕ, we also suppose that $\theta \mapsto \mathbb{E}(\phi(\theta, X))$ is a strictly convex function going to infinity at infinity and that a.s., the same property holds for $\theta \mapsto m_n(\theta, \phi)$ when n is large enough. A sufficient condition is the existence of a function $\varphi : \mathbb{R}_+ \mapsto \mathbb{R}$ such that $\lim_{t\to+\infty}\varphi(t) = +\infty$ and that

$$\mathbb{P}\left(\theta \mapsto \phi(\theta, X) \text{ strictly convex and not smaller than } \theta \mapsto \varphi(|\theta|)\right) > 0.$$

By adapting the proofs of Proposition 2.2 and Proposition 3.6, one obtains the following three results.

Proposition 4.1 *Assume that* $\mathbb{E}(\sup_{\theta\in K}|\phi(\theta, X)|) < \infty$ *for some compact neighborhood* K *of* $\theta_\star^\phi$. *Then, a.s.,* θ_n^ϕ *converges to* $\theta_\star^\phi$ *and* $m_n(\theta_\star^\phi, \phi)$ *converges to* $\mathbb{E}(\phi(\theta_\star^\phi, X))$.

Theorem 4.2 *Assume that the assumptions of Proposition 4.1 hold and that for a given compact neighborhood* K *of* $\theta_\star^\phi$, $\mathbb{E}(\sup_{\theta\in K}|h(\theta, X)|) < \infty$ *and a.s.,* $\theta \in K \mapsto h(\theta, X)$ *is continuous. Then, a.s.,* $m_n(\theta, h)$ *converges to* $\mathbb{E}(h(\theta, X))$ *uniformly on* K *and* $m_n(\theta_n^\phi, h)$ *converges to* $\mathbb{E}(h(\theta_\star^\phi, X))$.

Note that the a.s. convergence of $m_n(\theta_n, h)$ to $\mathbb{E}(h(\theta_\star^\phi, X))$ still holds for any sequence $(\theta_n)_n$ converging a.s. to $\theta_\star^\phi$.

Proposition 4.3 *Assume that the assumptions of Proposition 4.1 hold, that a.s., $\theta \in \mathbb{R}^d \mapsto \phi(\theta, X)$ is of class C^2 in the neighborhood of $\theta_\star^\phi$ and that $\mathbb{E}(|\nabla_\theta \phi(\theta_\star^\phi, X)|^2) < \infty$,*

$$\mathbb{E}\left(\sup_{\theta \in K} |\nabla_\theta \phi(\theta, X)| + \sup_{\theta \in K} |\nabla_\theta^2 \phi(\theta, X)|\right) < \infty$$

for some compact neighborhood K of $\theta_\star^\phi$. If the matrix $B = \mathbb{E}\left(\nabla_\theta^2 \phi(\theta_\star^\phi, X)\right)$ is non-singular, then $\sqrt{n}(\theta_n^\phi - \theta_\star^\phi) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, B^{-1}\mathrm{Cov}\,(\nabla_\theta \phi(\theta_\star^\phi, X))B^{-1})$.

Theorem 4.4 *Assume that the assumptions of Proposition 4.3 hold, that $\mathbb{E}(h^2(\theta_\star^\phi, X)) < +\infty$ and that for a given compact neighborhood K of $\theta_\star^\phi$, a.s., $\theta \mapsto h(\theta, \cdot)$ is C^1 on K and*

$$\mathbb{E}\left(\sup_{\theta \in K} |h(\theta, X)| + \sup_{\theta \in K} |\nabla_\theta h(\theta, X)|\right) < \infty.$$

Then, $\sqrt{n}(m_n(\theta_n^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X))) \xrightarrow{\mathcal{L}} \mathcal{N}_1\left(0, D^* \Sigma D\right)$ where

$$D = \begin{pmatrix} \mathbb{E}(\nabla_\theta h(\theta_\star^\phi, X)) \\ 1 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} B^{-1} & 0 \\ 0 & 1 \end{pmatrix} \mathrm{Cov} \begin{pmatrix} \nabla_\theta \phi(\theta_\star^\phi, X) \\ -h(\theta_\star^\phi, X) \end{pmatrix} \begin{pmatrix} B^{-1} & 0 \\ 0 & 1 \end{pmatrix}.$$

For this last result, we first prove that $\sqrt{n} \begin{pmatrix} \theta_n^\phi - \theta_\star^\phi \\ m_n(\theta_\star^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X)) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_{d+1}\left(0, \Sigma\right)$, using that the equality $\nabla_\theta m_n(\theta_n^\phi, \phi) = m_n(\theta_n^\phi, \nabla_\theta \phi) = 0 = \mathbb{E}\left(\nabla_\theta \phi(\theta_\star^\phi, X)\right)$ implies that

$$\left(\int_0^1 m_n(t\theta_n^\phi + (1-t)\theta_\phi^*, \nabla_\theta^2 \phi) dt\right)(\theta_n^\phi - \theta_\star^\phi) = \mathbb{E}\left(\nabla_\theta \phi(\theta_\star^\phi, X)\right) - m_n(\theta_\star^\phi, \nabla_\theta \phi),$$

where the first factor on the left-hand-side converges a.s. to B by Theorem 4.2. Then, the decomposition

$$\begin{aligned} \sqrt{n}(m_n(\theta_n^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X))) &= \sqrt{n}(m_n(\theta_n^\phi, h) - m_n(\theta_\star^\phi, h)) + \sqrt{n}(m_n(\theta_\star^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X))) \\ &= \sqrt{n}(\theta_n^\phi - \theta_\star^\phi) \cdot m_n(\bar{\theta}_n, \nabla_\theta h) + \sqrt{n}(m_n(\theta_\star^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X))) \end{aligned}$$

with $\bar{\theta}_n \in [\theta_n^\phi, \theta_\star^\phi]$ gives the result.

Example : In a financial context, [Avellaneda et al., 2001] propose a Monte-Carlo approach to transform an a priori probability measure μ on the space $\mathbb{R}^q$ of possible evolutions of the market into an a posteriori probability measure ν compatible with the market prices $C_1, \ldots, C_d$ of d financial assets defined by the measurable payoff functions $\bar{f}_1, \ldots, \bar{f}_d : \mathbb{R}^q \to \mathbb{R}$. For a sequence $(X_i)_{i \geq 1}$ of i.i.d. random vectors according to μ, these authors suggest to associate with the first n vectors $X_1, \ldots, X_n$ nonnegative weights $p_1^n, \ldots, p_n^n$ such that $\sum_{i=1}^n p_i^n = 1$ and $\forall k \in \{1, \ldots, d\}$, $\sum_{i=1}^n p_i^n \bar{f}_k(X_i) = C_j$. If the probability measure μ models the a priori knowledge of the market behavior, it is sensible to choose weights as close as possible to the uniform ones. For that purpose, [Avellaneda et al., 2001] propose to minimize $\sum_{i=1}^n p_i^n \log p_i^n$.

Let us define $f = (\bar{f}_1 - C_1, \ldots, \bar{f}_d - C_d)^*$. Assume from now on that

(H_1) : 0 belongs to the interior of the convex hull of the support of the image of μ by f,

and set $\phi(\theta, x) = e^{\theta.f(x)}$. Then, a.s. for n large enough 0 belongs to the interior of the convex hull of $\{f(X_1), \ldots, f(X_n)\}$, which ensures that $\theta \mapsto m_n(\theta, \phi)$ is a strictly convex function going to infinity at infinity. Note that in this context, each individual function $\theta \mapsto \phi(\theta, X_i)$ is convex but not strictly convex and does not go to infinity at infinity. There exists a unique $\theta_n^\phi \in \mathbb{R}^d$ such that $m_n(\theta_n^\phi, \phi) = \inf_{\theta \in \mathbb{R}^d} m_n(\theta, \phi)$. Introducing the Lagrangian associated with the strictly convex minimization problem with linear constraints as suggested by [Avellaneda et al., 2001], one easily checks that its solution is given by $p_i^n = e^{\theta_n^\phi . f(X_i)} / \sum_{j=1}^n e^{\theta_n^\phi . f(X_j)}$ for $i \in \{1, \ldots, n\}$. So, (H_1) is a sufficient condition for the existence of a solution for n large enough (see [Jourdain and Nguyen, 2001] for a more geometrically intricate necessary and sufficient condition). The price in the calibrated model of an exotic option with payoff function $g : \mathbb{R}^q \to \mathbb{R}$ is given by $m_n(\theta_n^\phi, \gamma)/m_n(\theta_n^\phi, \phi)$ where $\gamma(\theta, x) \stackrel{\text{def}}{=} g(x)e^{\theta.f(x)}$. The above theory enables to obtain the limiting behavior of both the optimal parameter θ_n^ϕ and the calibrated price. By (H_1), the function $\theta \mapsto \mathbb{E}(\phi(\theta, X_1)) = \mathbb{E}\left(e^{\theta.f(X_1)}\right)$ is strictly convex and goes to infinity at infinity. Since this function is lower semi-continuous by Fatou's Lemma and finite for $\theta = 0$, there exists a unique $\theta_\star^\phi$ such that $\mathbb{E}(\phi(\theta_\star^\phi, X_1)) = \inf_{\theta \in \mathbb{R}^d} \mathbb{E}(\phi(\theta, X_1))$. Under

(H_2) : $\theta_\star^\phi$ belongs to the interior of the convex set $\{\theta \in \mathbb{R}^d : \mathbb{E}\left(e^{\theta.f(X_1)}\right) < +\infty\}$,

since

$$\forall x \in \mathbb{R}^q, \quad \sup_{|\theta - \theta_\star^\phi| \leq \varepsilon} e^{\theta.f(x)} \leq \prod_{k=1}^d \left(e^{-\varepsilon f_k(x)} + e^{\varepsilon f_k(x)}\right) e^{\theta_\star^\phi . f(x)}.$$

According to Proposition 4.1, θ_n^ϕ and $m_n(\theta_n^\phi, \phi)$ respectively converge a.s. to $\theta_\star^\phi$ and $\mathbb{E}(\phi(\theta_\star^\phi, X_1))$. Note that according to [Jourdain and Nguyen, 2001], these convergence results still hold without assumption (H_2). Under (H_2) and

$$(H_3) : \exists \lambda, \beta > 0, \ \forall x \in \mathbb{R}^q, \ |g(x)| \leq \lambda(1 + |f(x)|^\beta),$$

Theorem 4.2 also implies that $m_n(\theta_n^\phi, \gamma)$ converges a.s. to $\mathbb{E}(\gamma(\theta_\star^\phi, X_1))$ so that the calibrated price converges a.s. to $\mathbb{E}\left(g(X_1)e^{\theta_\star^\phi . f(X_1)}\right) / \mathbb{E}\left(e^{\theta_\star^\phi . f(X_1)}\right)$. Under (H_2), the matrix $\mathbb{E}(\nabla_\theta^2 \phi(\theta_\star^\phi, X_1)) = \mathbb{E}\left(f(X_1)f(X_1)^* e^{\theta_\star^\phi . f(X_1)}\right)$ exists and is non-singular by (H_1). As a consequence, under (H_2) and

$$(H_4) : \mathbb{E}\left(|f(X_1)|^2 e^{2\theta_\star^\phi . f(X_1)}\right) < +\infty,$$

Proposition 4.3 gives the central limit theorem governing the convergence of θ_n^ϕ to $\theta_\star^\phi$. Note that a sufficient condition for (H_2) and (H_4) to hold is : $\forall \theta \in \mathbb{R}^d$, $\mathbb{E}\left(e^{\theta.f(X_1)}\right) < +\infty$. Last, since

$$\sqrt{n}\left(\frac{m_n(\theta_n^\phi, \gamma)}{m_n(\theta_n^\phi, \phi)} - \frac{\mathbb{E}\left(g(X_1)e^{\theta_\star^\phi . f(X_1)}\right)}{\mathbb{E}\left(e^{\theta_\star^\phi . f(X_1)}\right)}\right) = \frac{\sqrt{n}\left(m_n(\theta_n^\phi, h) - \mathbb{E}(h(\theta_\star^\phi, X_1))\right)}{m_n(\theta_n^\phi, \phi)\mathbb{E}\left(e^{\theta_\star^\phi . f(X_1)}\right)}$$

with $h(\theta,x) \stackrel{\text{def}}{=} \gamma(\theta,x)\mathbb{E}(\phi(\theta_\star^\phi,X_1))-\phi(\theta,x)\mathbb{E}(\gamma(\theta_\star^\phi,X_1))$, under (H_2), (H_3) and (H_4), by Slutsky's Theorem and Theorem 4.4, the left-hand-side converges in law to $\mathcal{N}_1\left(0, D^*\Sigma D/\mathbb{E}^4(e^{\theta_\star^\phi\cdot f(X_1)})\right)$ where the expressions of Σ and D are given in Theorem 4.4. Analogous central limit theorems are stated in a more general context in [Nguyen, 2003, chapter IV].

5 Practical implementation and applications

Option pricing in local or stochastic volatility models eventually boils down to the computation of an expectation $\mathbb{E}(f(G))$ where G is a $d-$dimensional standard normal random vector. In a financial context, there is no restriction in assuming that the payoff function f satisfies both (1.1) and (1.2). In most cases, this expectation will be computed using Monte Carlo simulations because closed formulas are barely available. The question of reducing the variance springs out quite naturally in this context. Relying on Equation (1.4), we have chosen the importance sampling point of view to tackle the delicate problem of variance reduction. Practitioners' concerns with variance reduction is to have an automatic toolbox at hand, which is precisely what we are devising here. As explained in the introduction, we advise to compute the minimizer $\vartheta_n^{f,A}$ of $v_n^{f,A}$ and then use this value in a Monte Carlo procedure as described in Algorithm 1. Note that the same samples are used to compute $\vartheta_n^{f,A}$ and the Monte Carlo estimator $M_n(A\vartheta_n^{f,A},f)$. Even though the terms involved in $M_n^f(A\vartheta_n^{f,A},f)$ are

Algorithm 1 Reduced Robust Importance Sampling (RRIS)

1. Generate $G_1,\ldots,G_n$ n i.i.d. samples following the law of G
2. Compute the minimizer $\vartheta_n^{f,A}$ of

$$v_n^{f,A}(\vartheta) = \frac{1}{n}\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i+\frac{|A\vartheta|^2}{2}}$$

3. Compute the expectation $\mathbb{E}(f(G))$ by Monte Carlo

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

not independent, according to Corollary 3.4, it is as easy to construct confidence intervals as for a crude Monte Carlo computation.

Remark 5.1 *In the name of the Algorithm 1, the term "Reduced" emphasizes that the optimal importance sampling parameter is searched for in a subspace of the set of all parameters. When the matrix $A = I_d$, the algorithm is simply denoted RIS because there is no dimension reduction anymore.*

In this section, we first explain how $\vartheta_n^{f,A}$ can be computed using Newton's optimization procedure. Then, we illustrate the efficiency of this robust variance reduction technique both in the multidimensional Black Scholes framework and in more general local volatility frameworks.

5.1 Solving the minimization problem

We already know from Proposition 2.1 that the function $v_n^{f,A}$ is strongly convex and infinitely continuously differentiable. Hence, we can approximate $\vartheta_n^{f,A}$ using Newton's algorithm for instance. The Hessian matrix $\nabla^2_\vartheta v_n^{f,A}(\vartheta)$ writes as the sum of a scalar matrix and a positive semi-definite matrix. Hence, it is quite obvious that the smallest eigenvalue of $\nabla^2_\vartheta v_n^{f,A}(\vartheta)$ is larger than the smallest eigenvalue of A^*A times $\frac{1}{n}\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i+\frac{|A\vartheta|^2}{2}}$. This last term can be arbitrary small depending on the function f. Therefore, implementing a straightforward Newton's algorithm can be particularly inefficient in some cases. It would be much better to have an alternative representation of $\vartheta_n^{f,A}$ as the minimizer of a function whose smallest eigenvalue of its Hessian matrix does not depend on f. We advise to rewrite $\nabla_\vartheta v_n^{f,A}(\vartheta)$ as

$$\nabla_\vartheta v_n^{f,A}(\vartheta) = A^*A\vartheta\,\frac{1}{n}\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i+\frac{|A\vartheta|^2}{2}} - \frac{1}{n}\sum_{i=1}^n A^*Gf^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i+\frac{|A\vartheta|^2}{2}}\,.$$

Hence, $\vartheta_n^{f,A}$ can be seen as the root of

$$\nabla_\vartheta u_n^{f,A}(\vartheta) = A^*A\vartheta - \frac{\sum_{i=1}^n A^*G_i f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}}{\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}},$$

with $u_n^{f,A}(\vartheta) = \frac{|A\vartheta|^2}{2} - \log\left(\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}\right)$. The Hessian matrix of $u_n^{f,A}$ is given by

$$\nabla^2_\vartheta u_n^{f,A}(\vartheta) = A^*A + \frac{\sum_{i=1}^n A^*G_iG_i^*Af^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}}{\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}} - \frac{\left(\sum_{i=1}^n A^*G_if^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}\right)\left(\sum_{i=1}^n A^*G_if^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}\right)^*}{\left(\sum_{i=1}^n f^2(G_i)\,\mathrm{e}^{-A\vartheta\cdot G_i}\right)^2}.$$

Using Cauchy-Schwartz's inequality, it is clear that $\nabla^2_\vartheta u_n^{f,A}(\vartheta) - A^*A$ is a positive semi-definite matrix. Hence, the smallest eigenvalue of $\nabla^2_\vartheta u_n^{f,A}(\vartheta)$ is always larger than the smallest one of A^*A whatever the values taken by f are. This advocates the use of $u_n^{f,A}$ rather than $v_n^{f,A}$ to compute $\vartheta_n^{f,A}$.

Using this new expression, we implement Algorithm 2 to construct an approximation x_n^k of $\vartheta_n^{f,A}$. Since $u_n^{f,A}$ is strongly convex, for any fixed n, x_n^k converges to $\vartheta_n^{f,A}$ when k goes to infinity. The direction of descent d_n^k at step k should be computed as the solution of a linear system. There is no in point in computing the inverse of $\nabla^2_\vartheta u_n^{f,A}(x_n^k)$, which would be computationally much more expensive.

Remarks on the implementation : From a practical point of view, ε should be chosen reasonably small $\varepsilon \approx 10^{-6}$. This algorithm converges very quickly and, in most cases, less than 5 iterations are enough to get a very accurate estimate of $\vartheta_n^{f,A}$, actually within the $\varepsilon-$error. Since the points at which the payoff function f is evaluated remain constant through the iterations of Newton's algorithm, the values $f^2(G_i)$ for $i = 1, \dots, n$ should be precomputed before starting the optimization algorithm which considerably speeds up the whole process. The Hessian matrix of our problem is easily tractable so there is no point in using Quasi-Newton's methods.

Algorithm 2 Newton's algorithm

Choose an initial value $x_n^0 \in \mathbb{R}^d$.
$k = 1$
while $|\nabla_\vartheta u_n^{f,A}(x_n^k)| > \varepsilon$ **do**
1. Compute d_n^k such that $(\nabla_\vartheta^2 u_n^{f,A}(x_n^k))d_n^k = -\nabla_\vartheta u_n^{f,A}(x_n^k)$
2. $x_n^{k+1} = x_n^k + d_n^k$, $k = k+1$

end while

5.2 Numerical examples

In this part, we present numerical results obtained by combining Algorithms 1 and 2 for different pricing problems. The variances for the RRIS algorithm (resp. the RIS algorithm) given in the tables below are computed along a single run of the algorithm using the estimator $v_n^{f,A}(\vartheta_n^{f,A}) - M_n^2(\theta_n^{f,A}, f)$ (resp. $v_n^f(\theta_n^f) - M_n^2(\theta_n^f, f)$) which converges almost surely to $v^f(\theta_\star^f) - \mathbb{E}^2(f(G))$ under the assumptions of Theorem 3.2. The variances of the crude Monte Carlo methods (denoted Var MC in the tables) are estimated by $\frac{1}{n}\sum_{i=1}^n f^2(G_i) - \left(\frac{1}{n}\sum_{i=1}^n f(G_i)\right)^2$. All the histograms presented hereafter are centered around their empirical means and renormalized by the empirical variances. When no further indications are given, the matrix A is chosen as the identity which implies that $d = d'$ and $\theta_n^{f,A} = \vartheta_n^{f,A}$.

5.2.1 Black-Scholes' framework

First, we consider an $I-$dimensional Black-Scholes model in which the dynamics under the risk neutral measure of each asset S^i is supposed to be given by

$$dS_t^i = S_t^i(rdt + \sigma^i dW_t^i) \qquad S_0 = (S_0^1, \dots, S_0^I)$$

where $W = (W^1, \dots, W^I)$. Each component W^i is a standard Brownian motion. For the numerical experiments, the covariance structure of W will be assumed to be given by $\langle W^i, W^j\rangle_t = \rho t \mathbf{1}_{\{i\neq j\}} + t\mathbf{1}_{\{i=j\}}$. We suppose that $\rho \in (-\frac{1}{d-1}, 1)$, which ensures that the matrix $C = (\mathbf{1}_{\{i\neq j\}} + \mathbf{1}_{\{i=j\}})_{1\leq i,j\leq I}$ is positive definite. Let L denote the lower triangular matrix involved in the Cholesky decomposition $C = LL^*$. To simulate W on the time-grid $0 < t_1 < t_2 < \ldots < t_N$, we need $d = I \times N$ independent standard normal variables and set

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

where G is a normal random vector in $\mathbb{R}^{I\times N}$. The vector $(\sigma^1, \dots, \sigma^d)$ is the vector of volatilities and $r > 0$ is the instantaneous interest rate. We will denote the maturity time by T.

Basket option We consider options with payoffs of the form $(\sum_{i=1}^d \omega^i S_T^i - K)_+$ where $(\omega^1, \dots, \omega^d)$ is a vector of algebraic weights. The strike value K can be taken negative to

deal with Put like options. All these payoffs belong to $\mathcal{H}_1$, so that Theorem 3.3 applies as Figures 1 and 2 illustrate it. These histograms have been obtained with 5000 independent runs of the RIS algorithm. The case of such basket options is definitely a burning issue because there is no closed formula as soon as $d > 2$ and the variance of a crude Monte Carlo approach can be dramatically large. We can see on the examples of the basket options treated

ρ	K	Price	Variance MC	Variance RIS
0.1	45	7.20	12.12	1.04
	55	0.56	1.90	0.14
0.2	50	3.29	13.56	1.74
0.5	45	7.68	42.2	5.06
	55	1.90	14.46	1.25
0.9	45	8.26	69.47	7.89
	55	2.84	30.08	2.58

Table 1: 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 $n = 10\,000$.

in Table 1 that the Robust Importance Sampling method does reduce the variance by at least 10. The results are obtained within 4.5 CPU seconds, compared to the 1.5 CPU seconds needed for the crude Monte Carlo computation. The same number of samples are used in both methods, which brings an overall gain of $10/\sqrt{3}$ in favor of the RIS algorithm. In the case $\rho = 0.2$ and $K = 50$, which is the option used for the histograms, the empirical variance is 1.76 whereas the on-line estimated variance is 1.74. This illustrates the conclusion of Corollary 3.4. The improvement brought by the RIS algorithm is very encouraging; not only because it definitely reduces the variance but above all because it is fully automatic. Unlike most adaptive importance sampling strategies developed so far and in particular the ones based on stochastic approximations, the one we propose here does not require any parameter tuning.

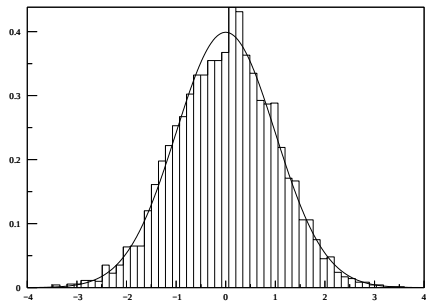

Figure 1: Limiting distribution of θ_n^f for the option of Table 1 with $\rho = 0.2$ and $K = 50$

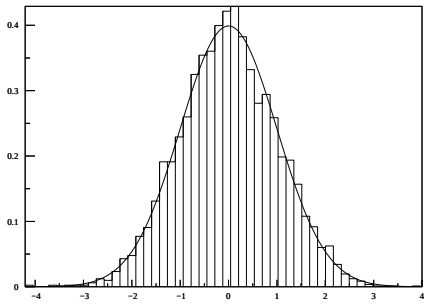

Figure 2: Limiting distribution of $M_n(\theta_n^f, f)$ (RIS) for the option of Table 1 with $\rho = 0.2$ and $K = 50$

We have also tested our algorithm on a 10−dimensional exchange option with randomly chosen spots and volatilities. The numerical results of Table 2 show that the RIS algorithm

performs well for a wide variety of basket options. In any case, the variance is at least divided 7, whereas it increases the CPU time by 4 times. This leads to an overall gain of 3.5 in the worst case.

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 2: Basket option in dimension $d = 10$ with $r = 0.05$, $T = 1$, $K = 0$, $\rho = 0.2$. The spots are chosen uniformly in $[70, 130]$ and the volatilities in $[0.1, 0.3]$. $\omega^i = \frac{1}{d}$ for $i = 1, \dots, d/2$ and $\omega^i = -\frac{1}{d}$ for $i = d/2 + 1, \dots, d$ and $n = 100\,000$.

One dimensional digital option We consider an option with payoff $\mathbf{1}_{\{S_T > L\}}$ where $L > 0$. We choose $T = 1$, $S_0 = 100$, $\sigma = 0.2$, $r = 0.05$ and $L = 140$. We fix the number of samples to 100 000. A crude Monte Carlo computation gives a price of 0.05952 with a variance of 0.053, whereas the exact price is 0.05968. On each run of the algorithm, we can compute the on-line estimator of the variance and use it to construct a confidence interval. We have run the RIS algorithm 100 000 times independently and on each run we have constructed the confidence interval of level 95% using the on-line estimated variance $v_n^{f,A}(\vartheta_n^{f,A}) - M_n^2(\theta_n^{f,A}, f)$. The true price falls outside the confidence interval in 5104 cases out of 100 000 which gives a level of 94.9%. This little experiment illustrates how Corollary 3.4 can be used to construct confidence intervals.

One dimensional barrier option This time, we only focus on one asset and we want to price a Call option with a discrete barrier on this asset. A discrete barrier means that we only check if the asset has crossed the barrier at fixed dates $t_1, \dots, t_d = T$, usually one per month. We assume that the grid defined by $t_1, \dots, t_d$ is regular with step size $\delta t = T/d$. The payoff can be written as $(S_T - K)_+ \mathbf{1}_{\{\forall 1 \le i \le d,\, S_{t_i} \ge L\}}$ for a Down and Out call option with barrier L. The price of such an option writes as $\mathbb{E}(f(G^1, \dots, G^d))$ with $f(x_1, \dots, x_d) = \mathrm{e}^{-rT}(S_0\, \mathrm{e}^{(r - \frac{\sigma^2}{2})T + \sigma\sqrt{\delta t}\sum_{j=1}^d x_j} - K)_+ \mathbf{1}_{\{\forall 1 \le i \le d,\, S_0\, \mathrm{e}^{(r - \frac{\sigma^2}{2})t_i + \sigma\sqrt{\delta t}\sum_{j=1}^i x_j} \ge L\}}$. In this particular case, if we consider the RIS algorithm developed before, the importance sampling parameter θ lies in $\mathbb{R}^d$. Hence, the optimization problem becomes harder to solve as the number of time steps increases.

One idea is to restrict the parameter θ to the subspace $\{A\vartheta : \vartheta \in \mathbb{R}\}$ where the vector A is defined by $A = (\sqrt{t_1}, \dots, \sqrt{t_d - t_{d-1}})^*$. In this case, the optimal parameter is always real-valued $d' = 1$ whatever the number of time steps we consider. This alternative approach — referred to as RRIS (Reduced Robust Importance Sampling) — corresponds to adding a linear drift to the Brownian motion. These two approaches are compared in Table 3 for the case of a Down and Out Call option and it turns out that the optimal variances obtained in both cases are very close to each other. When the underlying asset is of dimension one, the computation time gained by using the RRIS algorithm instead of the RIS one is not

that important but it will become a burning issue for multidimensional barrier options. The

L	Price	Variance MC	Variance RIS	Variance RRIS
70	11.46	401.51	34.10	34.33
80	11.19	401.04	35.68	36.11
90	9.60	383.93	42.54	45.37
95	7.56	342.05	42.01	49.84

Table 3: Down and Out Call option with $\sigma = 0.2$, $r = 0.05$, $T = 2$, $S_0^1 = 100$, $K = 110$ and $n = 10\,000$.

efficiency of the two algorithms on the Down and Out Call option is very impressive. Like in the previous example, the variance is reduced by a factor between 8 and 11. The use of the RRIS algorithm compared to a crude Monte Carlo method doubles the computation time, which means that the gain is at least $8/\sqrt{2}$. Figures 3 and 4 illustrate the asymptotic behavior of the RIS algorithm. They have been obtained by running the RIS algorithm 5 000 times independently. The histogram of Figure 3 represents the limiting distribution of the first component of θ_n^f computed with the RIS algorithm and rather well fits the density of the standard normal distribution (plain line), which illustrates Proposition 2.2. Although the hypotheses of Theorems 3.2 and 3.3 are not satisfied for the payoff at hand in the RIS framework, Figure 4 shows that our estimator is still convergent and asymptotically normal. This numerical convergence is emphasized by the matching of the empirical variance of the histogram and the on-line variance computed on a single run of the RIS algorithm; for these two quantities, we respectively find 34.70 and 35.68. Since the payoff belongs to $\mathcal{V}_A$, the convergence and the asymptotic normality of the RRIS estimator are in return ensured by Theorems 3.2 and 3.3.

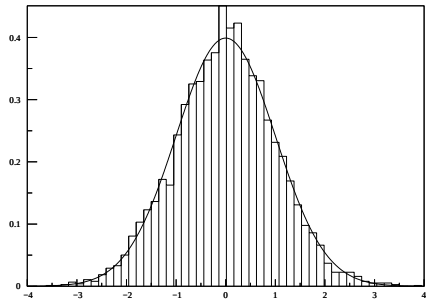

Figure 3: Limiting distribution of the first component of θ_n^f (RIS) for the option of Table 3 with $L = 80$

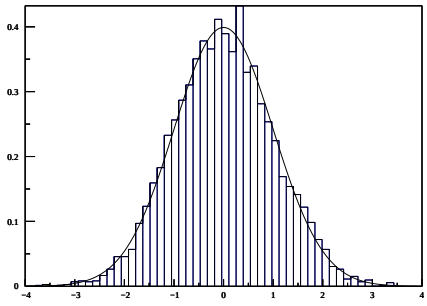

Figure 4: Limiting distribution of $M_n(\theta_n^f, f)$ (RIS) for the option of Table 3 with $L = 80$

Barrier Basket option We consider basket options in dimension I with a discrete barrier on each asset. For instance, if we consider a Down and Out Call option, the payoff writes down $(\sum_{i=1}^I \omega^i S_T^i - K)_+ \mathbf{1}_{\{\forall i \le I,\, \forall j \le N,\, S^i_{t_j} \ge L^i\}}$ where $\omega = (\omega^1, \ldots, \omega^I)$ is a vector of positive weights, $L = (L^1, \ldots, L^I)$ is the vector of barriers, $K > 0$ the strike value and $t_N = T$. Once

again, we consider one time step per month, which means that for an option with maturity time $T = 2$ as in Table 4, the number of time steps is $N = 24$. From now on, we fix $I = 5$. Hence, in the RIS algorithm the parameter θ is of dimension $d = 120$. Even though this is not that huge, it requires much more computational time as the numerical experiments show it. For the option of Table 4, a standard Monte Carlo computation takes 4.3 CPU seconds, the RRIS algorithm 8.7 CPU seconds whereas the RIS algorithm needs 22.5 CPU seconds. The RIS algorithm is three times slower than the RRIS algorithm in which the parameter θ lies in the subspace $\{A\vartheta : \vartheta \in \mathbb{R}^d\}$ of dimension $d' = I = 5$ with $A_{(j-1)I+i,i} = \sqrt{t_j - t_{j-1}}$ (convention $t_0 = 0$) for $j = 1, \ldots, N$ and $i = 1, \ldots, I$, all the other coefficients of A being zero.

Path dependent basket options are a prime case of pricing problems in which the use of one importance sampling parameter per time step dramatically slows down the computation. Restricting the importance sampling parameter space to a subspace of dimension $d' = I = 5$ as in the RRIS algorithm divides the computational time by 3, whereas the optimal variance of the RRIS algorithm is very close to the one of the RIS algorithm. Hence, there is no point in using one importance sampling parameter per time step. The improvement factor in terms of variance provided by the RRIS algorithm varies between 10 and 20. Because the RRIS algorithm is twice slower than a standard Monte Carlo computation, the overall gain factor varies between $10/\sqrt{2}$ and $20/\sqrt{2}$.

K	Price	Var MC	Var RIS	Var RRIS
45	2.37	22.46	2.58	2.62
50	1.17	10.97	0.78	0.79
55	0.51	4.72	0.19	0.19

Table 4: 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 $n = 100\,000$.

The payoff does not satisfy the assumptions of Theorems 3.2 and 3.3 neither in the RIS nor in the RRIS framework. Nevertheless, it looks pretty clear from Figure 6 that the RRIS estimator is convergent and asymptotically normal. Besides, for $K = 50$ the variance computed on a single run of the RRIS algorithm perfectly matches the empirical variance of the histogram. These histograms have been drawn with 100 000 independent runs of the RRIS algorithm.

5.2.2 Dupire's framework

We consider an $I-$dimensional local volatility model in which the dynamics under the risk neutral measure of each asset S^i is supposed to be given by

$$dS_t^i = S_t^i(rdt + \sigma(t, S_t^i)dW_t^i) \qquad S_0 = (S_0^1, \ldots, S_0^d)$$

where $W = (W^1, \ldots, W^I)$ is defined and generated like in the Black-Scholes' framework. The local volatility function σ we have chosen is of the form

$$\sigma(t, x) = 0.6(1.2 - \mathrm{e}^{-0.1t}\,\mathrm{e}^{-0.001(x\,\mathrm{e}^{rt} - s)^2})\,\mathrm{e}^{-0.05\sqrt{t}}, \tag{5.1}$$

with $s > 0$. We know that there exists a duality between the variables (t, x) and (T, K) in Dupire's framework. Hence for the formula (5.1) to make sense, one should choose s equal

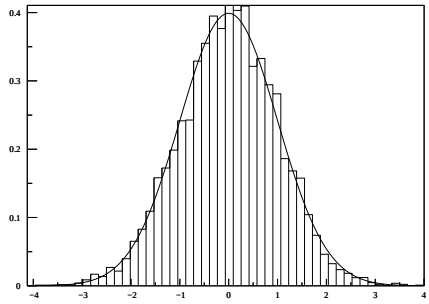

Figure 5: Limiting distribution of the first component of $\vartheta_n^{f,A}$ (RRIS) for the option of Table 4 with $K = 50$

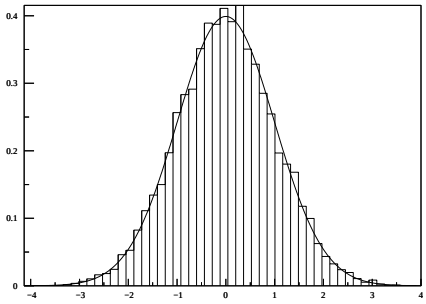

Figure 6: Limiting distribution of $M_n(\theta_n^{f,A}, f)$ (RRIS) for the option of Table 4 with $K = 50$

to the spot price of the undermying asset so that the bottom of the smile is located at the forward monney. We refer to Figure 7 to have an overview of the smile.

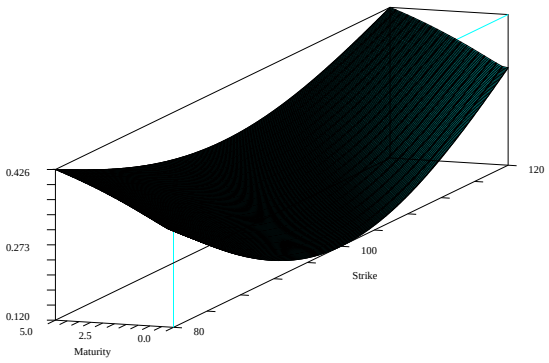

Figure 7: Local volatility function

Best Of option We consider options with payoffs $(\max_{1\leq i\leq I} \omega_i S_T^i - K)_+$, where $K > 0$ and $(\omega^1, \ldots, \omega^I)$ is a vector of positive weights. The payoffs belong to $\mathcal{H}_1$. To discretize the dynamics, we use an Euler scheme with $N = 100$ time steps per year. The results of Table 5 are encouraging. The RRIS algorithm with A defined like in the barrier basket option case reduces the variance by 4 whereas it only increases the computational time by 2 which leads to a gain of $4/\sqrt{2}$. We do not present any results for the RIS algorithm because the extra computational time it requires makes it noncompetitive.

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

Table 5: 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 $i = 1 \cdots d$.

Conclusion

We propose a fully automatic adaptive importance sampling technique for the computation of $\mathbb{E}(f(G))$ where $f : \mathbb{R}^d \to \mathbb{R}$ and G is a standard d-dimensional normal random vector. For a large class of functions f including many financial payoffs, we prove that our estimator is convergent and asymptotically normal with optimal limiting variance. Note that all the convergence results stated in Theorems 3.2, 3.3, Corollary 3.4, Propositions 3.7, 3.14, Lemma 3.8 and Remarks 3.10, 3.11, 3.16 still hold if $M_n(\theta, g)$ is defined as $\frac{1}{n}\sum_{i=1}^n g(\tilde{G}_i + \theta)e^{-\theta.\tilde{G}_i - \frac{|\theta|^2}{2}}$ for any sequence $(\tilde{G}_i)_{i\geq 1}$ of i.i.d. d-dimensional standard normal random vectors and in particular when this sequence is independent from the one $(G_i)_{i\geq 1}$ used to compute $(\vartheta_n^{f,A]})_{n\geq 1}$. Our numerical experiments confirm the performance of our estimator : in comparison with the crude Monte Carlo method, the computation time needed to achieve a given precision is divided by a factor going from 3 to 15. Moreover, they suggest that the convergence and asymptotic normality of the estimator still hold under weaker assumptions on the function f. In view of these numerical results and the definition of $\mathcal{V}_1$, it would be natural to investigate the class of functions f such that for some constants $\lambda > 0$ and $\beta \in [0, 2)$,

$$\forall \varphi : \mathbb{R}^d \to \mathbb{R}^d \; C^\infty \text{ and vanishing outside } B(0, M), \; \left| \int_{\mathbb{R}^d} f \nabla.\varphi(x) dx \right| \leq \lambda e^{M^\beta} \|\varphi\|_\infty.$$

Unfortunately, we have not been able so far to derive the asymptotic properties of our estimator for such functions. In this work, we have focused on importance sampling. A natural extension would be to investigate the coulping with statification techniques in the spirit of [Glasserman et al., 1999]. In particular, it would be interesting to combine the present importance sampling algorithm with the adaptive stratified sampling methods proposed recently in [Etoré and Jourdain, 2008] (adaptive optimization of the proportions of random drawings made in the different strata) and [Etoré et al., 2008] (adaptive optimization of the stratification direction $e \in \mathbb{R}^d$ for a standard normal random vector when the strata are given by $\{x \in \mathbb{R}^d : e \cdot x \in [y_{i-1}, y_i)\}$ with $-\infty = y_0 < y_1 < y_2 < \ldots < y_I = +\infty$).

References

[Arouna, 2004] Arouna, B. (2004). Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24.

[Arouna, 0304] Arouna, B. (Winter 2003/04). Robbins Monro algorithms and variance reduction in finance. *J. of Computational Finance*, 7(2).

[Avellaneda et al., 2001] Avellaneda, M., Buff, R., Friedman, C., Grandchamp, N., Kruk, L., and Newman, J. (2001). Weighted Monte Carlo: a new technique for calibrating asset-pricing models. *Int. J. Theor. Appl. Finance*, 4(1):91–119.

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

[Chen et al., 1988] Chen, H. F., Lei, G., and Gao, A. J. (1988). Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2):217–231.

[Etoré et al., 2008] Etoré, P., Fort, G., Jourdain, B., and Moulines, E. (2008). On adaptive stratification. Technical report, hal-00319157.

[Etoré and Jourdain, 2008] Etoré, P. and Jourdain, B. (2008). Adaptive optimal allocation in stratified sampling methods. *Methodol. Comput. Appl. Probab.* To appear.

[Glasserman, 2004] Glasserman, P. (2004). *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York)*. Springer-Verlag, New York. , Stochastic Modelling and Applied Probability.

[Glasserman et al., 1999] Glasserman, P., Heidelberger, P., and Shahabuddin, P. (1999). Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance*, 9(2):117–152.

[Jourdain and Nguyen, 2001] Jourdain, B. and Nguyen, L. (2001). Minimisation de l'entropie relative par méthode de Monte-Carlo. *C. R. Acad. Sci. Paris Sér. I Math.*, 332(4):345–350.

[Kim and Henderson, 2004] Kim, S. and Henderson, S. G. (2004). Adaptive control variates. In *Proceedings of the 2004 Winter Simulation Conference*.

[Kim and Henderson, 2007] Kim, S. and Henderson, S. G. (2007). Adaptive control variates for finite-horizon simulation. *Math. Oper. Res.*, 32(3):508–527.

[Lavenberg et al., 1982] Lavenberg, S. S., Moeller, T. L., and Welsh, P. D. (1982). Statistical results on control variables with application to queueing network simulation. *Operations Research*, 30(1):182 – 202.

[Ledoux and Talagrand, 1991] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.

[Lelong, 2007] Lelong, J. (2007). *Asymptotic properties of stochastic algorithms and pricing of Parisian options*. PhD thesis, Ecole Nationale des Ponts et Chaussées.

[Lelong, 2008] Lelong, J. (2008). Almost sure convergence of randomly truncated stochastic algorithms under verifiable conditions. *Statistics & Probability Letters*, 78(16).

[Lemaire and Pagès, 2008] Lemaire, V. and Pagès, G. (2008). Unconstrained Recursive Importance Sampling. Technical report, arXiv 0807.0762.

[Nelson, 1990] Nelson, B. L. (1990). Control variate remedies. *Oper. Res.*, 38(6):974–992.

[Nguyen, 2003] Nguyen, L. (2003). *Calibration de Modèles Financiers par Minimisation d'Entropie Relative et Modèles avec Sauts.* PhD thesis, Ecole Nationale des Ponts et Chaussées.

[Rubinstein and Shapiro, 1993] Rubinstein, R. Y. and Shapiro, A. (1993). *Discrete event systems.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester. Sensitivity analysis and stochastic optimization by the score function method.