Robbins-Monro algorithms and variance reduction

Bouhari AROUNA
CERMICS, Ecole Nationale des Ponts et Chaussées, 6 et 8 av Blaise Pascal,
77455 Marne La Vallée, France, email : arouna@cermics.enpc.fr

September, 10 2002

Abstract :

In this paper, we present a new variance reduction technique for Monte Carlo methods. By an elementary version of Girsanov theorem, we introduce a drift term in a price computation. Afterwards, the basic idea is to use a truncated version of the Robbins-Monro (RM) algorithms to find the optimal drift that reduces the variance. We proved that for a large class of payoff functions, this version of RM algorithms converges *a.s.* to the optimal drift. It is also shown that an adaptative use of this RM algorithm inside the Monte Carlo computation leads to an effective variance reduction. In this scope, we find some limit theorems which enable us to illustrate the method by applications to options pricing.

1 Introduction

Monte carlo methods are used for pricing and hedging complex financial products especially when the number of the assets involved is large. In such a case, variance reduction methods are often needed in order to improve efficiency. In this paper we present importance sampling methods based on Girsanov transformation following [6]. The basic idea is to use a Robbins-Monro (RM) algorithm to optimize the choice of the drift in the Girsanov transformation. The RM algorithm is a stochastic approximation method which allows to estimate asymptotically the zeros of a function given as an expectation. Although its rate of convergence is $C/\sqrt{n}$ in general, the RM algorithm is very easy to implement in general. Newton [13] proved that for a large class of problems of options pricing in continuous time, importance sampling can lead to a zero-variance estimator through a *stochastic* change of drift. However, determining the optimal drift requires knowing the option's price in advance. This approach is therefore based on using approximations of the option's price to find approximations of the optimal drift. We use a

different approach: we restrict ourselves to deterministic change of drift.
In the next section we present the mathematical context of our method and introduce briefly the importance sampling technique based on Girsanov transformation (following [6]). In section 3, we first introduce the RM algorithms in a general framework and then present the Chen's method which enable us to prove our main result. Section 4 deals with some relevant limit theorems for the algorithm. The last part of the work is devoted to numerical tests and practical considerations . A brief presentation of the RM algorithm using Chen's truncation method is given in the appendix (see [7]).

2 Mathematical Context

2.1 Financial background

Let us assume that the price of the underlying asset under the risk neutral probability is described by the stochastic differential equation

$$dS_t = S_t(rdt + \sigma(t, S_t)dW_t),\ S_0 = x, \tag{1}$$

with r the risk-free, continuously compounded interest rate, $\sigma(t, y)$ the asset's volatility, W_t a brownian motion, and x fixed. By arbitrage, the price of an option with payoff $\psi(S_t, t \leq T)$ is given by

$$V_0 = \mathbb{E}[e^{-rT}\psi(S_t, t \leq T)]. \tag{2}$$

For practical purposes, we will restrict attention to simulations of the asset driven by a sequence of independent normal variables, since we can recover this case when the normal variables are correlated through a linear transformation. However, we have in mind simulations through discritizations of diffusion processes using for example, an *Euler Scheme*, when an exact solution of the stochastic differential equation (1) is not available. We assume that an acceptable discretization of this equation has already been determined on a dicrete grid of points $0 = T_0 < T_1 < \cdots < T_m = T$, and thus we focus attention on obtaining precise estimates of the price V_0. Therefore, in a practical situation, to compute V_0, we have to evaluate

$$\hat{V}_0 = \mathbb{E}[e^{-rT}\psi(S_{T_1}, \ldots, S_{T_m})],$$

which we rewrite as

$$\hat{V}_0 = \mathbb{E}[G(Z)], \tag{3}$$

where $Z = (Z_1, \ldots, Z_m) \sim \mathcal{N}(0, I_m)$ and G is a function we can compute using the dicretization of S .
In what follows, the objective is to evaluate (3) using an importance sampling procedure.

2.2 Importance sampling

We change the law of $Z = (Z_1, \ldots, Z_m)$ adding a drift vector $\mu = (\mu_1, \ldots, \mu_m)$. An elementary version of Girsanov theorem -see for example [2] or [4] for details- applied to (3) leads to the following representation of $\hat{V}_0$:

$$\hat{V}_0 = \mathbb{E}(\alpha(\mu)) \ , \tag{4}$$

with

$$\alpha(\mu) = G(Z+\mu)e^{(-\mu\cdot Z - \frac{1}{2}\|\mu\|^2)} \ , \tag{5}$$

where $\|x\|$ denotes the Euclidean norm of a vector $x \in \mathbb{R}^m$. The authors in [6] give an importance sampling procedure to minimize the variance of $\alpha(\mu)$ or equivalently to minimize $\mathbb{E}(\alpha^2(\mu))$ with respect to μ. This method reduces the contribution of the linear part of the "log-payoff" to the variance by sampling along a direction $\hat{\mu}$ which is solution to the fixed-point problem : $\nabla \log G(\mu) = \mu$.
In this paper, we use a RM algorithm to assess the "optimal sampling direction" μ^* that minimizes the variance of $\alpha(\mu)$, $\mu \in \mathbb{R}^m$ or equivalently :

$$H(\mu) = \mathbb{E}(\alpha^2(\mu)). \tag{6}$$

For more convenient we write $g(\mu, z)$ for the value of $G(z+\mu)e^{-\mu\cdot z - \frac{1}{2}\|\mu\|^2}$ and $s_p(\mu)$ for the value of $\mathbb{E}(g^p(\mu, Z))$. The following result is important.

Proposition 2.1. *If $\mathbb{E}(G^{2a}(Z)) < \infty$, with $a > 1$, then H is twice differentiable in $\mathbb{R}^m$ and there exists a unique $\mu^* \in \mathbb{R}^m$ such that :*

$$H(\mu^*) = \min_{\mu \in \mathbb{R}^m} H(\mu). \tag{7}$$

Proof Using Girsanov theorem, we obtain

$$H(\mu) = \mathbb{E}\Big[G^2(Z)e^{-\mu\cdot Z + \frac{1}{2}\|\mu\|^2}\Big]. \tag{8}$$

Suppose that $\|\mu\| \leq K$ where K is a non negative constant. With the notation $h(\mu, z) = (\mu - z)G^2(z)e^{-\mu\cdot z + \frac{1}{2}\|\mu\|^2}$, we have

$$\int |h(\mu, z)|e^{-\frac{1}{2}\|z\|^2}dz \leq e^{\frac{K^2}{2}} \int (K + \|z\|)e^{K\|z\|}e^{-(1-\frac{1}{a})\frac{1}{2}\|z\|^2}G^2(z)e^{-\frac{1}{2a}\|z\|^2}dz.$$

By Hölder's inequality, we can write

$$\int |h(\mu, z)|e^{-\frac{1}{2}\|z\|^2}dz \leq e^{\frac{K^2}{2}} \left(\int (K + \|z\|)e^{\frac{aK}{a-1}\|z\|}e^{-\frac{1}{2}\|z\|^2}dz\right)^{1-\frac{1}{a}} \times$$
$$\times \left(\int G^{2a}(z)e^{-\frac{1}{2}\|z\|^2}dz\right)^{\frac{1}{a}}.$$

Since $\mathbb{E}(G^{2a}(Z)) < \infty$, it is not difficult to see that H is differentiable and that

$$\nabla H(\mu) = \mathbb{E}\Big[(\mu - Z)G^2(Z)e^{-\mu\cdot Z + \frac{1}{2}\|\mu\|^2}\Big]. \qquad (*)$$

In addition, one can prove that H is twice differentiable and that

$$\mathrm{Hess}H(\mu) = \mathbb{E}\Big[\Big(I_m + (\mu - Z)(\mu - Z)^T\Big)G^2(Z)e^{-\mu\cdot Z + \frac{1}{2}\|\mu\|^2}\Big], \qquad (**)$$

where $\mathrm{Hess}H(\cdot)$ denotes the hessian matrix of H and I_m the identity matrix of size m. From $(**)$, we conclude that H is strictly convex on $\mathbb{R}^m$ since $\forall\ u \in \mathbb{R}^m - \{0\}$,

$$u^T \mathrm{Hess}H(\mu)u = \mathbb{E}\Big[\Big(\|u\|^2 + (u\cdot(\mu - Z))^2\Big)G^2(Z)e^{-\mu\cdot Z + \frac{1}{2}\|\mu\|^2}\Big] > 0.$$

To end this proof, it's sufficient to show that $\lim_{\|\mu\|\to+\infty} H(\mu) = +\infty$. Using Jensen inequality, it follows that

$$\begin{aligned}\log H(\mu) &\geq \mathbb{E}\Big(2\log G(Z)\mathbf{1}_{\{G>0\}} - \mu Z + \frac{1}{2}\|\mu\|^2\Big)\\ &= 2\mathbb{E}(\log G(Z)\mathbf{1}_{\{G>0\}}) + \frac{1}{2}\|\mu\|^2.\end{aligned}$$

Therefore, if $\mathbb{P}(G(Z) > 0) \neq 0$, then $\lim_{\|\mu\|\to+\infty} H(\mu) = +\infty$. $\Box$
As a consequence of the proposition above, μ^* minimizing H is the unique solution of

$$\nabla H(\mu) = 0\ , \qquad (9)$$

and the idea is to make use of a RM algorithm to solve equation (9).

3 Robbins-Monro algorithms

We begin this section by a short presentation of the Robbins-Monro algorithms. Afterwards we introduce the Chen's truncation method.

3.1 General framework

The RM algorithms have the form

$$X_{n+1} = X_n - \gamma_{n+1}F(X_n, Z_{n+1}) \qquad (10)$$

where Z_n is drawn from a given distribution $\mathbf{m}(dx)$.
The initial condition is any admissible value for X_0. This algorithm solves the equation

$$\mathbb{E}[F(\mu, Z)] = 0$$

where $\mathbb{E}$ denotes the expectation under $\mathbf{m}(dx)$. If we consider the mean field

$$h(\mu) = \mathbb{E}[F(\mu, Z)], \quad \mu \in \mathbb{R}^m,$$

we can rewrite (10) as

$$X_{n+1} = X_n - \gamma_{n+1} h(X_n) + \gamma_{n+1} \epsilon_{n+1} \tag{11}$$

with

$$\epsilon_{n+1} = h(X_n) - F(X_n, Z_{n+1}).$$

The ϵ_n can be seen as random errors made when evaluating $h(X_n)$. Let us write Y_{n+1} for the value of $F(X_n, Z_{n+1})$. X_n and Y_n are random vectors in $\mathbb{R}^m$. Let $\mathcal{F}_n = \sigma\{X_k,\ Y_k,\ k \leq n\}$ be the σ-algebra generated by $X_k,\ Y_k$ $fork \leq n$. Clearly we can write

$$\mathbb{E}[Y_{n+1}/\mathcal{F}_n] = h(X_n).$$

The following theorem is proved in [10] or [9].

Theorem 1. *Under the following hypothesis*

$$(H_1) \quad \exists \mu^* \in \mathbb{R}^m,\ h(\mu^*) = 0, \quad \forall \mu \in \mathbb{R}^m\ \mu \neq \mu^*\ (\mu - \mu^*) \cdot h(\mu) > 0, \tag{12}$$

$$(H_2) \qquad \sum_n \gamma_n = +\infty \quad \textit{and} \quad \sum_n \gamma_n^2 < +\infty, \tag{13}$$

$$(H_3) \qquad \mathbb{E}[\|Y_{n+1}\|^2/\mathcal{F}_n] < K(1 + \|X_n\|^2)\ a.s., \tag{14}$$

the sequence of random vectors $(X_n)_{n\geq 0}$ converges almost surely to μ^*.

One can find some other convergence hypothesis of the RM algorithms in [14], [1], [10] or in [11] for a simple presentation of stochastic algorithms. Some papers are devoted to the convergence properties of these algorithms see e.g. [5] and [8]. Unfortunately, classical theorems such as the one above can not be used in the case we are concerned with.

3.2 Application to variance reduction

In our case (see $(*)$), the mean field h is given by

$$h(x) = \mathbb{E}\left[(x - Z) G^2(Z) e^{-x \cdot Z + \frac{1}{2}\|x\|^2}\right], \tag{15}$$

where Z is dawn from the gaussian law $\mathcal{N}(0, I_m)$. By *Proposition 2.1*, it exists a unique $\mu^* \in \mathbb{R}^m$ which makes zero the function h. Now, consider the following expression of Y_{n+1} :

$$Y_{n+1} = (X_n - Z_{n+1})G^2(Z_{n+1})e^{-X_n \cdot Z_{n+1} + \frac{1}{2}\|X_n\|^2}, \tag{16}$$

where $(Z_n)_{n\geq 0}$ is a sequence of *i.i.d.* gaussian vectors following the law of Z. Since X_n is $\mathcal{F}_n$-mesurable and Z_{n+1} is independent of $\mathcal{F}_n$, it is easy to see that

$$\mathbb{E}[Y_{n+1}/\mathcal{F}_n] = h(X_n).$$

Hypothesis (H_1) of the theorem above is satisfied by h and hypothesis (H_2) is a question of trivial choice. On the contrary, hypothesis (H_3) can not be satisfied. Obviously this fact is due to the exponential form of Y_{n+1}. Hence the most difficult point to check is that X_n does not tend to infinity. To deal with this particular point, we use a technique introduced by H.F. Chen in [8] (see also [7]) using projections to get convergence.

3.3 Truncation method

To describe the method, first fix $x^1 \neq x^2$ in $\mathbb{R}^m$ and choose a constant $M > 0$ as indicated in the appendix. Let $(Z_n)_{n\geq 0}$ be a sequence of independent random vectors drawn from the distribution of Z. Let $(U_n)_{n\geq 0}$ be an arbitrary deterministic increasing sequence of positive numbers tending to infinity with $U_0 > M$.
Define for $n \geq 0$,

$$X_{n+1} = \begin{cases} X_n - \gamma_{n+1}Y_{n+1} & \text{if} \quad \|X_n - \gamma_{n+1}Y_{n+1}\| \leq U_{\sigma(n)}, \\ x_n^* & \text{otherwise} \end{cases} \tag{17}$$

$$\sigma(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\|X_k - \gamma_{k+1}Y_{k+1}\| > U_{\sigma(k)}}, \quad \sigma(0) = 0, \tag{18}$$

$\sigma(n)$ is the number of projections done after n iterations.

$$x_n^* = \begin{cases} x^1 & \text{if } \sigma(n) \text{ is even,} \\ x^2 & \text{if } \sigma(n) \text{ is odd,} \end{cases} \tag{19}$$

with $(\gamma_n)_{n\geq 0}$ a sequence of positive numbers satisfying

$$\sum_{n\geq 0} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 0} \gamma_n^2 < +\infty. \tag{20}$$

Remark 3.1. In our numerical tests we use $\gamma_n = \frac{\alpha}{\beta+n}$, $\alpha, \beta > 0$. The problem of the "best choice" of the coefficients α and β is rather delicate. From a numerical point of view, this choice seems to be linked to the values of the model parameters. However, we propose in the last section an empirical and intuitive way of choosing efficiently these coefficients.

Remark 3.2. The constant M above seems to have no significant effect on the numerical convergence of the algorithm when took a reasonable value. In our numerical tests M values are in the range $[10, 100]$ with no effect on the convergence properties of the algorithm. At time n, x_n^* may be a function of the past values of the algorithm. For example a randomly chosen former points.

The following lemma allows us to apply the result of Chen to our settings.

Lemma 1. *1) It exists a twice continuously differentiable function $v : \mathbb{R}^m \to \mathbb{R}$, such that :*

$$v(x^*) = 0, \quad \lim_{\|x\| \to \infty} v(x) = +\infty$$

$$\text{and } \forall \quad x \neq x^* \quad v(x) > 0, \quad h(x) \cdot \nabla v(x) > 0.$$

2) Let G satisfies $\mathbb{E}(|G(Z)|^{4p}) < +\infty$ with $p > 1$, then one can choose the sequence U_n such that

$$\lim_{n \to +\infty} \sum_{k \leq n} \gamma_{k+1}^2 \mathbb{E}[\|Y_{k+1}\|^2 / \mathcal{F}_k] < +\infty \quad a.s.$$

Proof Let $v(x) = \|x - x^*\|^2$. By *Propositon 2.1*, the function H defined by

$$H(x) = \mathbb{E}[G^2(Z) e^{-x \cdot Z + \frac{1}{2}\|x\|^2}]$$

is strictly convex and its gradient is given by

$$h = \mathbb{E}[(x - Z) G^2(Z) e^{-x \cdot Z + \frac{1}{2}\|x\|^2}].$$

Thus

$$\forall \quad u \neq y \in \mathbb{R}^m \quad H(y) - H(u) > (y - u) \cdot h(u),$$

and for $y = \mu^*$ we have

$$\forall \quad u \neq \mu^* \quad H(u) - H(\mu^*) < \nabla v(u) \cdot h(u) \ .$$

As $\forall \quad u \in \mathbb{R}^m$, $H(\mu^*) < H(u)$, the first part of the lemma is proved. To prove the last part, first observe that X_n is $\mathcal{F}_n$-mesurable and Z_{n+1} is independent of $\mathcal{F}_n$. Then we have

$$\mathbb{E}[\|Y_{n+1}\|^2 / \mathcal{F}_n] = s^2(X_n)$$

with

$$s^2(x) = \mathbb{E}[\|x - Z\|^2 G^4(Z) e^{-2x\cdot Z + \|x\|^2}].$$

Now write $h(x,z)$ for the value of $\|x - Z\|^2 e^{-2x\cdot Z+\|x\|^2}$. $\forall \beta > 1$ we have

$$\begin{aligned}
\mathbb{E}(h^{\beta}(x,Z)) &= \mathbb{E}(\|x - Z\|^{2\beta} e^{-2\beta x\cdot Z + \beta\|x\|^2}) \\
&= \mathbb{E}(\|Z\|^{2\beta} e^{(2\beta+1)x\cdot Z - (\beta+\frac{1}{2})\|x\|^2}) \\
&\le e^{-(\beta+\frac{1}{2})\|x\|^2} \mathbb{E}(\|Z\|^{2\beta} e^{(2\beta+1)\|x\|\|Z\|}) \\
&= e^{-(\beta+\frac{1}{2})\|x\|^2} \int_{\mathbb{R}^+\times(0,2\pi)\times(0,\pi)^{m-2}} r^{2\beta} e^{(2\beta+1)\|x\| r} e^{-\frac{r^2}{2}} J_{\phi}(r,\theta_{m-1},\ldots,\theta_1)\, dr\, d\theta_{m-1}\ldots d\theta_1
\end{aligned}$$

where $J_{\phi}(r,\theta_{m-1},\ldots,\theta_1) = r^{m-1} f(\theta_{m-1},\ldots,\theta_1)$ denotes the Jacobian of the transformation from Cartesian to polar variables. Therefore

$$\begin{aligned}
\mathbb{E}(h^{\beta}(x,Z)) &\le C_1(\beta,m) e^{-(\beta+\frac{1}{2})\|x\|^2} \int_0^{+\infty} r^{2\beta+m-1} e^{(2\beta+1)\|x\| r} e^{-\frac{r^2}{2}} dr \\
&\le C_1(\beta,m) e^{-(\beta+\frac{1}{2})\|x\|^2} \int_0^{+\infty} r^{2\beta+m-1} e^{(2\beta+1)^2\|x\|^2 + \frac{r^2}{4}} e^{-\frac{r^2}{2}} dr \\
&= C_1(\beta,m) e^{-(\beta+\frac{1}{2})\|x\|^2 + (2\beta+1)^2\|x\|^2} \int_0^{+\infty} r^{2\beta+m-1} e^{-\frac{r^2}{4}} dr \\
&= C_2(\beta,m) e^{-(\beta+\frac{1}{2})\|x\|^2 + (2\beta+1)^2\|x\|^2}.
\end{aligned}$$

Now by the Hölder inequality, it follows

$$\begin{aligned}
s^2(x) &\le \left(\mathbb{E}(|G(Z)|^{4p})\right)^{\frac{1}{4p}} \left(\mathbb{E}(h^{\frac{4p}{4p-1}}(x,Z))\right)^{1-\frac{1}{4p}} \\
&\le C(p,m) e^{Q(p)\|x\|^2}.
\end{aligned}$$

where $C_1(\beta,m)$, $C_2(\beta,m)$ and $C(p,m)$ are three positive constants, and $Q(p)$ is defined by $Q(p) := \frac{7.5p^2 - p + 1/32}{p^2 - p/4}$. Using equations (17-20), we get

$$\|X_n\| \le \max(U_{\sigma(n)}, \|x_n^*\|) \le U_n, \quad \text{and} \quad s^2(X_n) \le C e^{Q(p)U_n^2},$$

for n sufficiently large. Finally we conclude the proof by choosing the sequence U_n such that

$$\sum_n \gamma_n^2 e^{Q(p)U_n^2} < +\infty . \square$$

Remark 3.3. The sequence U_n must increase sufficiently slow to cancel the explosion behaviour of the algorithm without modifications on it and the choice of U_n is not difficult. In fact $\forall p > 1$ the sequence $U_n = \sqrt{\frac{1}{10}\ln n} + U_0$, $n \ge 1$ is suitable since $Q(]1,+\infty[) \subset]7.5, 8.71[$.

The following theorem is a consequence of the lemma above.

Theorem 2. *In the framework of Lemma 1, the algorithm X_n defined by (17-20) converges a.s. to the unique solution of the equation $h(x) = 0$, $x \in \mathbb{R}^m$ and the number of truncations $\sigma(n)$ is bounded.*

Proof First, set $\epsilon_{n+1} = h(X_n) - Y_{n+1}, \quad n \geq 0$ and define the sequence $M_n = \sum_{i=0}^{n-1} \gamma_{i+1}\epsilon_{i+1}$ for $n \geq 1$ and $M_0 = 0$. The sequence $(M_n)_{n\geq 1}$ is a $\mathcal{F}_n$-martingale and its brackets process is given by

$$\begin{aligned}
\langle M \rangle_n &= \sum_{i=0}^{n-1} \gamma_{i+1}^2 \mathbb{E}\Big[\|\epsilon_{i+1}\|^2/\mathcal{F}_i\Big] \\
&= \sum_{i=0}^{n-1} \gamma_{i+1}^2 \mathbb{E}[\|Y_{i+1}\|^2/\mathcal{F}_i] - \sum_{i=0}^{n-1} \gamma_{i+1}^2 \|h(X_i)\|^2 \\
&\leq \sum_{i=0}^{n-1} \gamma_{i+1}^2 \mathbb{E}[\|Y_{i+1}\|^2/\mathcal{F}_i].
\end{aligned}$$

Using *Lemma 1*, we have choosen the sequence (U_n) such that

$$\begin{aligned}
\lim_{n\to+\infty} \langle M \rangle_n &\leq \sum_{n=0}^{+\infty} \gamma_{n+1}^2 \mathbb{E}[\|Y_{n+1}\|^2/\mathcal{F}_n] \quad a.s. \\
&\leq C \sum_n \gamma_n^2 e^{Q(p)U_n^2} \quad a.s. \\
&< +\infty \quad a.s.,
\end{aligned}$$

where $C > 0$. Therefore the martingale M_n converges $a.s.$ and in $\mathcal{L}^2$ (see [12] or [3]). The Kronecker's lemma (see for example [15] p.117) implies that $\overline{\lim}\, \gamma_{n+1} \| \sum_{i=0}^{n-1} \epsilon_{i+1} \| = 0 \quad a.s.$. Chen, Guo and Gao proved in [7] that assuming the first part of *Lemma 1* holds, one just need the additional assumption $\overline{\lim}\, \gamma_{n+1} \| \sum_{i=0}^{n-1} \epsilon_{i+1} \| = 0 \quad a.s.$ in order to obtain the convergence of the algorithm. *Theorem 2* is then a consequence of their result which is stated in *Theorem 6* (see appendix or [7]).□

In practical situations, one don't need to know the exact value of the optimal drift vector μ^*. As good the convergence of the RM algorithm is towards μ^*, as good is the variance reduction obtained. The algorithm above may be summarize as the following,

(a) First use the Robbins Monro algorithm X_n defined by (17-20) to assess μ^*,

(b) Then, inject the value of μ^* in the Monte Carlo method by computing $\hat{V}_0 \simeq \frac{1}{N} \sum_{n=1}^{N} g(\mu^*, Z_n)$.

We call this method : "algorithm1".

We show in the sequel that it is possible to merge parts (a) and (b) of above, by computing directly

$$\hat{V}_0 \simeq \frac{1}{N} \sum_{n=1}^{N} g(X_{n-1}, Z_n).$$

This last computation simplify "algorithm1"; we denoted it by "algorithm2".

4 Law of Large Numbers and Central Limit Theorem

We need the following classical results for the remainder.

Lemma 2. *If f is a continuous function from $\mathbb{R}^d$ to $\mathbb{R}$ and $(x_n)_{n\geq 0}$ is a sequence of real numbers which converges to x, then*

$$\frac{1}{n} \sum_{k=1}^{n} f(x_k) \longrightarrow f(x), \text{ as } n \to +\infty.$$

Theorem 3. *Let $(M_n)_{n\geq 0}$ be a real, square-integrable martingale which is adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$ and has an bracket process denoted by $\langle\ M\ \rangle_n$. Suppose that for a real deterministic sequence $(a_n)_{n\geq 0}$ increasing to $+\infty$ the following two assumptions apply:*

(A) $\dfrac{\langle\ M\ \rangle_n}{a_n} \xrightarrow{\mathbb{P}} \sigma^2 \quad (\sigma > 0)$;

(B) Lindberg's condition holds; in order words, for all $\epsilon > 0$,
$\frac{1}{a_n} \sum_{k=1}^{n} \mathbb{E}\Big[\|M_k - M_{k-1}\|^2 1_{\{\|M_k - M_{k-1}\| \geq \epsilon\sqrt{a_n}\}} / \mathcal{F}_{k-1} \Big] \xrightarrow{\mathbb{P}} 0.$

Then: $\dfrac{M_n}{a_n} \xrightarrow{a.s.} 0$ *and* $\dfrac{M_n}{\sqrt{a_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Remark 4.1. The proof of the lemma is rather trivial.The theorem is proved in various books which give it its due importance (see [12] for example). The proof of the result $\dfrac{M_n}{a_n} \xrightarrow{a.s.} 0$ which represents the "Law of Large Numbers" part of this theorem lies only on assumption (A). The Lindberg's condition is essential to prove the last part of this theorem.

We will also use the following lemma.

Lemma 3. *If $\mathbb{E}(|G^{ap}(Z)|) < \infty$, with $a > 1$ and $p \geq 1$, then s_p is a continuous function from $\mathbb{R}^d$ to $\mathbb{R}$.*

Proof By Girsanov theorem it is rather trivial to see that

$$s_p(\mu) = \mathbb{E}(G^p(Z)e^{-(p-1)\mu\cdot Z+\frac{p-1}{2}\|\mu\|^2}).$$

Suppose that $\|\mu\| \leq K$ where K is a non negative constant. With the notation $h_p(\mu, z) = G^p(z)e^{-(p-1)\mu\cdot z+\frac{p-1}{2}\|\mu\|^2}$, we have

$$|h_p(\mu, z)| \leq e^{\frac{K^2}{2}(p-1)}|G^p(z)|e^{K(p-1)\|z\|}.$$

Using Hölder's inequality, it follows

$$\int |G^p(z)|e^{K(p-1)\|z\|}e^{-\frac{1}{2}\|z\|^2}dz \leq \left(\int e^{\frac{aK}{a-1}(p-1)\|z\|}e^{-\frac{1}{2}\|z\|^2}dz\right)^{1-\frac{1}{a}}\left(\int |G^{ap}(z)| \times e^{-\frac{1}{2}\|z\|^2}dz\right)^{\frac{1}{a}}.$$

Since $\mathbb{E}(|G^{pa}(Z)|) < \infty$, the Lebesgue theorem applies and the function s_p is continuous. □

Theorems 4 and 5 below may be seen as our main results. They lead to an effective variance reduction algorithm.

Theorem 4. *Assume that the Robbins-Monro algorithm $(X_n)_{n\geq 0}$ defined by (17-20) converges a.s. to μ^* and that for $p \geq 4$, $\mathbb{E}(|G^{ap}(Z)|) < \infty$, with $a > 1$. If $\overline{X}_N = \frac{1}{N}\sum_{n=1}^{N} g(X_{n-1}, Z_n)$ then :*

(1) $\overline{X}_N \xrightarrow{a.s.} \hat{V}_0$;

(2) $\sqrt{N}(\overline{X}_N - \hat{V}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, *with* $\sigma^2 = Var(g(\mu^*, Z))$.

Remark 4.2. We amphasize that $\sigma^2 = Var(g(\mu^*, Z))$ is the smallest variance one could expect with this "finite-dimensional" importance sampling method. As usual, for practical situations we need an asymptotical estimation of σ. This is the fact of the theorem below.

Theorem 5. *In the framework of the theorem above, if we write $\overline{\sigma}_N^2$ for the value of $\frac{1}{N}\sum_{n=1}^{N} g^2(X_{n-1}, Z_n) - \overline{X}_N^2$, then :*

(1) $\overline{\sigma}_N^2 \xrightarrow{a.s.} \sigma^2$;

(2) $$\frac{\sqrt{N}(\overline{X}_N - \hat{V}_0)}{\overline{\sigma}_N} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Remark 4.3. The payoff of most of products traded on financial markets satisfies the integrability condition of the theorems.

Proof of Theorem 4. Let us write M_N for the value of $\sum_{n=1}^{N}(g(X_{n-1}, Z_n) - \hat{V}_0)$. Since X_{n-1} is $\mathcal{F}_{n-1}$ mesurable and Z_n is independent of $\mathcal{F}_{n-1}$, by the virtue of (4-5), we have

$$\mathbb{E}(g(X_{n-1}, Z_n)/\mathcal{F}_{n-1}) = \hat{V}_0$$

and $(M_N)_{N\geq 0}$ is a martingale. Throughout the proof of *Lemma 3* and the definition of X_n, it appears that $(M_N)_{N\geq 0}$ is powered-p integrable, with $p \geq 4$. It's bracket process is given by

$$\begin{aligned}
\langle M \rangle_N &= \sum_{n=1}^{N} \mathbb{E}\Big(|g(X_{n-1}, Z_n) - \hat{V}_0|^2/\mathcal{F}_{n-1} \Big) \\
&= \sum_{n=1}^{N} \mathbb{E}\Big(g^2(X_{n-1}, Z_n)/\mathcal{F}_{n-1} \Big) - N\hat{V}_0^2 \\
&= \sum_{n=1}^{N} s_2(X_{n-1}) - N\hat{V}_0^2 \quad \text{with} \quad s_2(x) = \mathbb{E}(g^2(x, Z)).
\end{aligned}$$

By *Lemma 3* $s_2(x)$ is continuous and by *Lemma 2*

$$\frac{1}{N}\sum_{n=1}^{N} s_2(X_{n-1}) \xrightarrow{a.s.} s_2(\mu^*).$$

Therefore

$$\frac{\langle M \rangle_N}{N} \xrightarrow[N]{a.s.} \sigma^2, \quad \text{where} \quad \sigma^2 = s_2(\mu^*) - \hat{V}_0^2 = Var(g(\mu^*, Z)).$$

According to the part (A) of *Theorem 3*, we have $\dfrac{M_N}{N} \xrightarrow[N]{a.s.} 0$ which is equivalent to the part (1) of the theorem.
It remains to prove the Lindberg's condition in order to get the second part of the theorem. First we observe that

$$\begin{aligned}
\mathbb{E}\Big[|g(X_{n-1}, Z_n) - \hat{V}_0|^4/\mathcal{F}_{n-1} \Big] &= \mathbb{E}[g^4(X_{n-1}, Z_n)/\mathcal{F}_{n-1}] - 3\hat{V}_0^4 \\
&\quad - 4\hat{V}_0 \mathbb{E}[g^3(X_{n-1}, Z_n)/\mathcal{F}_{n-1}] \\
&\quad + 6\hat{V}_0^2 \mathbb{E}[g^2(X_{n-1}, Z_n)/\mathcal{F}_{n-1}] \\
&= s_4(X_{n-1}) - 3\hat{V}_0^4 - 4\hat{V}_0 s_3(X_{n-1}) \\
&\quad + 6\hat{V}_0^2 s_2(X_{n-1}).
\end{aligned}$$

Using again *Lemma 3* and *Lemma 2* we have

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\Big[|g(X_{n-1}, Z_n) - \hat{V}_0|^4/\mathcal{F}_{n-1} \Big] \xrightarrow{a.s.} L$$

where $L < +\infty \quad a.s.$ is a positive random variable. Now for $A > 0$ define

$$F_N(A) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[|g(X_{n-1},Z_n) - \hat{V}_0|^2 1_{\{|g(X_{n-1},Z_n)-\hat{V}_0|>A\}}/\mathcal{F}_{n-1}\Big].$$

It is easily seen that

$$F_N(A) \le \frac{A^{-2}}{N}\sum_{n=1}^{N}\mathbb{E}\Big[|g(X_{n-1},Z_n) - \hat{V}_0|^4/\mathcal{F}_{n-1}\Big],$$

so that

$$\limsup_{N\to+\infty} F_N(A) \le A^{-2}L \quad a.s.$$

Hence taking $A_N = \epsilon\sqrt{N}$ with $\epsilon > 0$, we have

$$\limsup_{N\to+\infty} F_N(\epsilon\sqrt{N}) = 0 \quad a.s.$$

and the Lindberg's condition holds. Finaly, *Theorem 3.* shows that

$$\frac{M_N}{\sqrt{N}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2) \quad \text{with} \quad \sigma^2 = Var(g(\mu^*,Z))$$

which is the desired conclusion.□

Proof of Theorem 5. To prove the first part of the theorem, we only need to show that

$$\frac{1}{N}\sum_{n=1}^{N} s_2(X_{n-1}) \sim \frac{1}{N}\sum_{n=1}^{N} g^2(X_{n-1},Z_n), \quad N \sim +\infty.$$

In this scope, let us denote $\overline{M}_N = \sum_{n=1}^{N}\Big(g^2(X_{n-1},Z_n) - s_2(X_{n-1})\Big)$. Obviously $\mathbb{E}(g^2(X_{n-1},Z_n)/\mathcal{F}_{n-1}) = s_2(X_{n-1})$ $a.s.$ so that $\overline{M}_N$ is a martingale. $\overline{M}_N$ is squared-integrable and has the following bracket process

$$\begin{aligned}\langle\overline{M}\rangle_N &= \sum_{n=1}^{N}\mathbb{E}\Big((g^2(X_{n-1},Z_n) - s_2(X_{n-1}))^2/\mathcal{F}_{n-1}\Big)\\ &= \sum_{n=1}^{N}\Big(s_4(X_{n-1}) - s_2^2(X_{n-1})\Big).\end{aligned}$$

Again by combining *Lemma 3* and *Lemma 2* it appears that

$$\frac{\langle\overline{M}\rangle_N}{N} \xrightarrow[N]{a.s.} s_4(\mu^*) - s_2^2(\mu^*) = Var(g^2(\mu^*,Z)) > 0.$$

Then *Theorem 3* shows once again that

$$\frac{1}{N}\sum_{n=1}^{N}\Big(g^2(X_{n-1},Z_n)-s_2(X_{n-1})\Big)\xrightarrow[N]{a.s.}0.$$

The second part of the theorem is a classical result of Probability theory.□

Remark 4.4. Exactly as in the *i.i.d.* case, *Theorem 4.* and *Theorem 5.* show that $\overline{X}_N=\frac{1}{N}\sum_{n=1}^{N}g(X_{n-1},Z_n)$ converges almost surely to the desired expectation $\hat{V}_0=\mathbb{E}[G(Z)]$ and that the rate of convergence is $\dfrac{1}{\sqrt{N}}$.

5 Examples and numerical tests

As noticed in *Remark 3.1* the "best" choice of the steps sequence $(\gamma_n)_{n\geq 0}$ in the algorithm (17-20) is rather delicate. From a theoretical point of view it is known (see [9] or [14]) that the best sequence must decrease towards 0 as $\frac{1}{n}$. In our numerical tests we use $\gamma_n=\frac{\alpha}{\beta+n}$, $\alpha,\beta>0$. We observe that the choice of β has no significant effect on the numerical convergence of the algorithm. The most difficult point to check for numerical purposes is therefore to find the values of the parameter α which lead to good convergence properties. We have represented the ratio of the classical Monte Carlo estimator's standard deviation to the one of the Monte Carlo method with the optimal drift computed by the method we proposed. We denoted this ratio by "StdRatio". Figure 1 shows the StdRatio obtained for a european

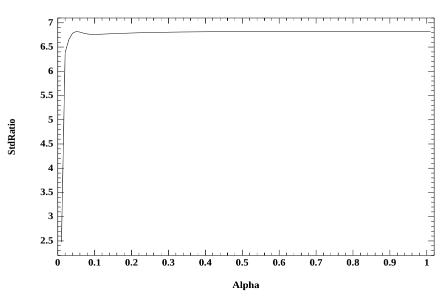

Figure 1: Variance reduction for a european call with "algorithm1"

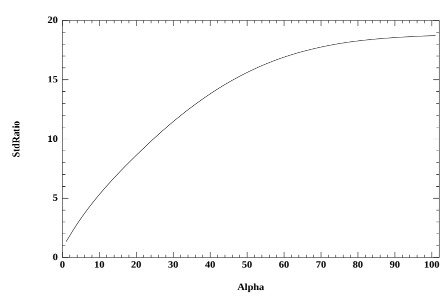

Figure 2: Variance reduction for a european put with "algorithm1"

call when α varies. This option is out-of-the money and its parameters are $S_0=50$, $K=80$, $r=0.05$, $T=1.0$, and $\sigma=0.3$. Figure 2 represents the StdRatio for a european put that is out of-the-money. In this case we use $S_0=50$, $K=40$, $r=0.05$, $T=1.0$, $\sigma=0.1$. In Figure 3 and Figure 4, we plot the StdRatio in a multidimensional case, namely the arithmetic asian put's case. We use $n=40$ discretization steps for the left panel and $n=20$ for the right one. The left panel represents this StdRatio when the put is out

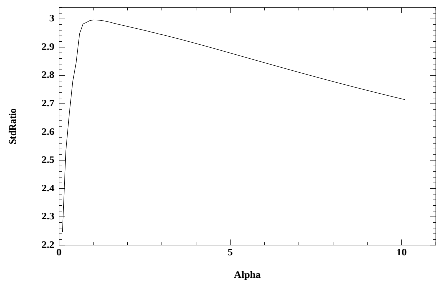

Figure 3: Variance reduction for an out of the money asian put -n=40- with “algorithm1”

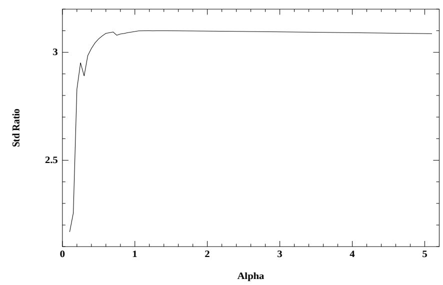

Figure 4: Variance reduction for an at the money asian put -n=20- with “algorithm1”

of the money with $S_0 = 50$, $K = 45$, $r = 0.05$, $T = 1.0$, and $\sigma = 0.1$. The right one shows the StdRatio for the same put, but at the money. Figure 5

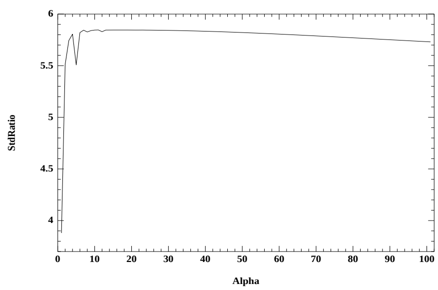

Figure 5: Variance reduction for a european basket put with “algorithm1”

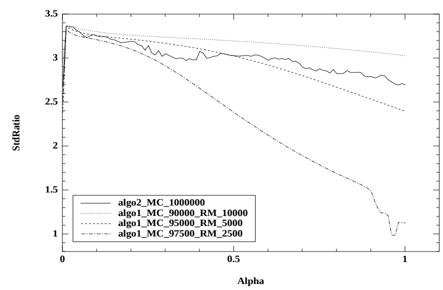

Figure 6: Variance reduction for a at-the-money european basket call

displays the StdRatio variations with respect to α, for a out of the money european put on a basket of $n = 10$ assets. We have used in this case the parameters $K = 30$, $r = 0.05$, $S_0 = 50$, $\sigma = 0.1$, and $T = 1$.

We begin the presentation of the results obtained by a one dimensional option pricing problem. Tables 4.1 and 4.2 present these results for european standard call and put. Of course the pricing of these products is available in closed form, but it seems natural for us to start the numerical tests with simple examples in order to measure both gain on variance and accuracy on prices computation.

Figure 6 deals with some comparisons between “algorithm1” and “algorithm2”. All the results are based on a total of 1,000,000 runs. *algo2_MC_1000000* illustrate an adaptative use of the RM algorithm in a Monte Carlo computation and presents the “StdRatio” reached -“algorithm2”.

algo1_MC_90000_RM_10000 , *algo1_MC_95000_RM_5000,* and *algo1_MC_97500_RM_2500* provides some aspects of the use of “algorithm1”. In such cases, we use respectively $90,000$, $95,000$, and $97,500$ simulated paths in the Monte Carlo computations, when $10,000$, $5,000$, and $2,500$ are respectively the number of RM algorithm's iterations. From this example, it seems that $2,500$ runs are not sufficient to guarantee a good convergence of the RM algorithm. In this particular case, the sensitivity of the algorithm with respect to α -Alpha in the figure- is higher than in the others.

Table 4.1

Estimated Variance Reduction Ratio for the European Put using "algorithm1"

Parameters				Importance sampling			
α	β	σ	strike	RMPrice	BSPrice	CPrice	StdRatio
5.	1.	0.3	30	0.13	0.13	0.13	6.2
0.1			40	1.28	1.28	1.27	3.3
0.01			50	4.68	4.68	4.70	2.5
0.001			60	10.54	10.53	10.58	2.2
100.		0.1	40	0.0042	0.0042	0.0040	18.7
1.			50	0.97	0.96	0.96	3.1
0.1			60	7.31	7.31	7.33	2.5

All the results are based on a total of 50,000 runs. 40,000 runs for the Monte Carlo method and 10,000 runs for the RM algorithm. The model parameters are: $S_0 = 50$, $r = 0.05$, and $T = 1.0$.

Tableau 4.2

Estimated Variance Reduction Ratio for the European Call using "algorithm1"

Parameters				Importance sampling			
α	β	σ	strike	RMPrice	BSPrice	CPrice	StdRatio
0.01	1.	0.3	30	21.63	21.60	21.52	4.1
0.1			50	7.12	7.12	7.01	3.3
0.5			60	3.45	3.45	3.43	3.9
0.1			80	0.67	0.67	0.68	6.8
0.0006		0.1	30	21.47	21.46	21.52	10.6
0.01			50	3.41	3.40	3.38	2.8
0.07			60	0.23	0.23	0.23	5.6
5.			70	0.004	0.004	0.004	25.

All the results are based on a total of 50,000 runs. 40,000 runs for the Monte Carlo method and 10,000 runs for the RM algorithm. The model parameters are $S_0 = 50$, $r = 0.05$, and $T = 1.0$.

"RMPrice", "CPrice" and "BSPrice" denote respectively the Monte Carlo estimated price including our method (Monte Carlo + Importance sampling + RM algorithm), the classical Monte Carlo price and the Black and Scholes exact price of the option. We recall that "StdRatio" is the ratio of the classical Monte Carlo estimator standard deviation to the one of the Monte Carlo using the optimal drift computed by our method.

On these simple examples the standard deviation error reduction is very significant. For a put and a call that are out of the money, the gain factor (StdRatio) could be high. Furthermore the prices computed by the Monte Carlo method including the variance reduction method we propose are very accurate. Table 4.3 and 4.4 show the variance reduction obtained with our method in the case of european basket call and put. The results are interesting, since the reduction of confidence interval length is about a factor of at least 2. This gain factor may be "large" for options that are out of the money.

Tableau 4.3

Estimated Variance Reduction Ratio for the European basket call using "algorithm2"

Parameters					Importance Sampling	
n	α	β	σ	strike	RMPrice	StdRatio
10	0.01	1.	0.1	40	13.41	3.3
				50	7.42	3.4
				60	3.76	3.8
	0.001		0.2	40	17.86	4.2
				50	13.36	4.3
				60	10.03	4.5
20	0.001		0.1	40	15.14	3.6
	0.01			50	9.89	3.7
				60	6.33	4.1
	0.001		0.2	40	21.80	5.4
				50	18.06	5.4
				60	15.12	5.6

We use 1,000,000 simulated paths for n=10 and 2,000,000 paths for n=20. The option parameters are $S_0 = 50$, $r = 0.05$, and $T = 1.0$,.

The numerical cost of this method is equivalent to the additional time spent in generating the gaussian paths that are used to compute the optimal drift. In all our tests this extra time does not exceed 20% of the CPU time spent in the classical Monte Carlo computation. In fact we use at most 20% gaussian paths in addition to those simulated for the standard Monte Carlo computation. The variance is reduced by a factor of at least 4. This reduction has reached a factor of 625 in our examples. Obviously, this gain justify the extra effort of computation. In table 4.4 we use respectively 900,000 and 100,000 simulation paths for Monte Carlo computation and Robbins Monro algorithm. In this particular case only 10% additional simulation effort leads to a variance reduction with a factor of at least 4.

Tableau 4.4

Estimated Variance Reduction Ratio for the European

Basket Put using "algorithm1"

Parameters					Importance sampling	
n	α	β	σ	strike	RMPrice	StdRatio
10	100	1.	0.1	20	0.003	14.9
	4.			30	0.18	5.8
	0.1			40	1.47	3.2
	0.01			50	5.00	2.5
	0.01			60	10.86	2.2
20	5.	1.		20	0.074	7.1
	0.5			30	0.84	3.7
	0.05			40	3.19	2.7
	0.01			50	7.48	2.3
	0.01			60	13.43	2.1
10	1.	1.	0.2	20	0.52	4.0
	0.05			30	2.39	2.8
	0.01			40	5.93	2.3
	0.01			50	10.95	2.1
	0.002			60	17.14	2.0

The number of assets involved is n. All the results use a total of 1,000,000 gaussian paths including 100,000 paths for the drift computation. The model parameters are $S_0 = 50$, $r = 0.05$, and $T = 1.0$. Volatility is flat at 10% or 20%.

Table 4.5 displays values of an arithmetic asian put. As one can notice, the variance gain is greater than a factor of 4. It is well known that put options variance is comparatively lower than call one since put payoff is bounded. Then a variance reduction with a factor of 4 is not negligible in the case fo a put.

The next example we considered deals with the Heston (1993) stochastic volatility model given by,

$$dS_t = rS_t dt + \sqrt{v_t} S_t dW_t^1,$$
$$dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t^2,$$

where W^1 and W^2 are two correlated brownian motions with $\langle W^1, W^2\rangle_t = \rho t$, and k, θ and σ are constants. Discretizing with an *Euler scheme* leads to

$$S_{T_{i+1}} = S_{T_i}(1 + r\Delta t + \sqrt{\sigma_i \Delta t} Z_i),$$
$$v_{T_{i+1}} = v_{T_i} + k(\theta - v_{T_i})\Delta t + \sigma\sqrt{\Delta t v_{T_i}}(\rho Z_i + \sqrt{1-\rho^2} Z_{m+i}),$$

Where $(Z_i)_{i\geq 1}$ is a sequence of independent gaussian variables with mean 0 and variance 1.

Tableau 4.5

Estimated Variance Reduction Ratio for the Asian Put using "algorithm1"

Parameters					Importance Sampling	
n	α	β	σ	strike	RMPrice	StdRatio
20	5	1.	0.1	45	0.013	5.8
	1			50	0.63	3.1
	0.05			55	3.68	2.5
20	6	1.	0.3	40	0.27	4.8
	0.5			50	2.87	2.6
	0.05			60	9.30	2.2
40	5	1.	0.1	45	0.011	4.3
	1			50	0.62	3.0
	0.05			55	3.70	2.4
40	4.5	1.	0.3	40	0.25	4.4
	1.			50	2.83	2.6
	0.05			60	9.29	2.2

We use 1,000,000 paths for the Monte Carlo computation and 200,000 for the optimal drift computation. The option parameters value are $S_0 = 50$, $r = 0.05$, and $T = 1.0$,.

Tableau 4.6

Estimated Variance Reduction Ratio for the European call in the Heston stochastic volatility model with "algorithm2".

Parameters				Importance sampling		
v_0	α	β	strike	ClosPR	RMPrice	StdRatio
0.01	0.001	1.	45	7.27	7.25	1.6
	0.01		50	3.33	3.33	2.4
	0.1		55	1.13	1.15	2.9
	1.0		60	0.32	0.33	3.4
0.04	0.001		45	7.67	7.60	1.9
	0.005		50	4.27	4.25	2.5
	0.2		55	2.11	2.10	3.1
	0.04		60	0.95	0.95	3.4

The number of discretization steps is n=100. All the results use a total of 2,000,000 gaussian paths. The model parameters are $S_0 = 50$, $r = 0.05$, $k = 2.0$, $\theta = 0.01$, $\rho = 0.5$, $T = 1.0$ and the volatility of the volatility is $\sigma = 0.1$.

Using this model, Heston has given a closed form solution to the pricing of a european call option by the caracteristic functions technique. We implement this closed form solution and in table 4.6, "ClosPR" denotes the price of the call computed with this solution. Table 4.6 dispalys also the pricing results obtained with our method ("algorithm2"). It appears that our method applies again in this case. One notice that notice that variance reduction can reach a factor of 12. Our last examples deal with a stochastic volatility model, namely the Hull-White stochastic volatility model (1987),

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{\sigma_t} S_t dW_t^1, \\ d\sigma_t &= \nu \sigma_t dt + \zeta \sigma_t dW_t^2, \end{aligned}$$

where W^1 and W^2 are two correlated brownian motions with $\langle W^1, W^2 \rangle_t = \rho t$. In this model, S_t has a finite mean but an infinite variance. Using a linear discretization of S_t by an *Euler scheme*, the variance is finite but increases very quickly with the number of steps. To reduce this effect, we need to truncate this variance. As in [6] we consider the following discretisation of the model

$$\begin{aligned} S_{T_{i+1}} &= S_{T_i}(1 + r\Delta t + \sqrt{\sigma_i \Delta t} Z_i), \\ \sigma_{i+1} &= \min\{c, \sigma_i e^{(\nu - \frac{1}{2}\zeta^2)\Delta t + \zeta\sqrt{\Delta t}(\rho Z_i + \sqrt{1-\rho^2} Z_{m+i})}\}, \end{aligned}$$

where c is a non-negative constant. The truncation has little impact on the mean but makes estimated variances much more stable.

Through our simulation results we take $c = 2$, $\nu = 0$, $r = 0.05$, $S_0 = 50$, $T = 1$, $\rho = 0.5$ and $\sqrt{\sigma_0} = 0.1$. The constant volatility case corresponds to $\zeta = 0$. The implementation of the method is not more difficult than it was in the Black and Scholes model. Again we plot the ratio of the classical Monte Carlo method's standard deviation error to that of Monte Carlo using our variance reduction method with respect to various α using "algorithm1". The payoff we consider is again that of a put option on the arithmetic mean

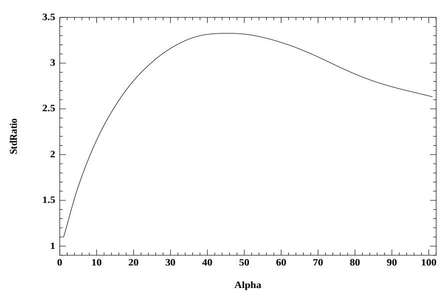

Figure 7: Out of the money asian put. Hull-White stochastic volatility model. $\zeta = 0.5$, $K = 45$

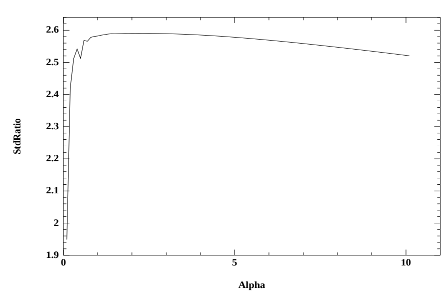

Figure 8: At the money asian put. Hull-White stochastic volatility model. $\zeta = 0.5$, $K = 50$

$$\hat{S} = \frac{1}{m} \sum_{i=1}^{m} S_{T_i}.$$

Results are based on a total of 1,000,000 paths for the Monte Carlo computation and a total of 200,000 for the optimal drift computation when using "algortihm1". Again, we can see through these examples that the confidence interval length reduction is greater than a factor of 2.

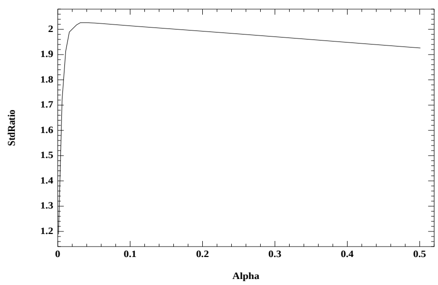

Figure 9: Out of the money asian put. Hull-White stochastic volatility model. $\zeta = 1,\ K = 45$

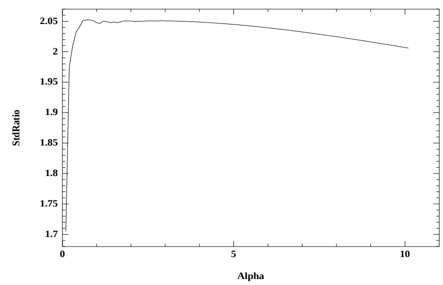

Figure 10: At the money asian put. Hull-White stochastic volatility model. $\zeta = 1,\ K = 50$

6 Concluding remarks

The method we propose in this paper is very general. It can be used as soon as a Monte Carlo method is feasible and it is very easy to implement.
It does not require regularity conditions on the payoff function. It could work both for path-dependent and path-independent products. In high dimensional problems, instead of choosing the steps sequence parameters arbitrarily, one can use the same simulation paths to compute both prices and variances with respect to these parameters. The price which corresponds to the smallest variance should give the best Monte Carlo estimation of the real price needed. To the best of our knowledge, the use of Robbins Monro algorithms in a Monte Carlo procedure in order to reduce variance is new.
The method proposed here could naturally be improved. For example, in order to get more stability an efficiency of the algorithm, one may use random vector fields which has lower variance. If F_1 and F_2 are two random vector fields -and h the mean field- in the algorithm (17-20) such that:

$$\mathbb{E}(F_1(X_n, Z_{n+1})/\mathcal{F}_n) = \mathbb{E}(F_2(X_n, Z_{n+1})/\mathcal{F}_n) = h(X_n),$$

then the one which has the lowest conditional variance seems to be the best.
To end these remarks, we make some simple and useful observations : in a price computation using Monte Carlo method, one might guess whether the Monte Carlo variance would be large or not. When the variance is likely to take large values, the parameter α should be small (ex. $\alpha \sim 0.001$ or 0.01). At the opposite, if the variance is likely to be small, α should be relatively

large (about 10 or 100).

References

[1] A. Benveniste, M. Metivier, and P. Priouret. *Algorithmes adaptatifs et approximations stochastiques.* Masson, 1987.

[2] B. Lapeyre, A. Sulem, and D. Talay. *Understanding Numerical Analysis for Financial Models.* "outcoming", France, 2002.

[3] D. D.-Castelle and D. Marie. *Probabilités et Statisques (tome 2.).* Masson, France, 1983.

[4] D. Lamberton and B. Lapeyre. *Introduction au calcul stochastique appliqué à le finance.* ellipse, 1997.

[5] D.S. Clark . Necessary and sufficient conditions for the robbins-monro method. *Stochastic Processes and their Applications*, 17:359–367, 1984.

[6] Glasserman P., Heidelberger P., and Shahabuddin P. Asymptotic optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9:117–152.

[7] H.-F. Chen, L. Guo, and A.-J. Gao. Convergence and robustness of the robbins-monro algorithm truncated at randomly varying bounds. *Stochastic Processes and their Applications*, 27:217–231, 1988.

[8] H.F. Chen and Y.M. Zhu. Stochastic approximation procedure with randomly varying truncations. *Scientia Sinica (series A)*, 29.

[9] D. Marie. *Algorithmes stochastiques.* Springer-Verlag, France, 1996.

[10] D. Marie. *Random Iterative Models.* Springer-Verlag, France, 1997.

[11] N. Martoli and P. Del Moral. *Simulation et algorithmes stochastiques.* Cépaduès-éditions, 2002.

[12] J. Neveu. *Martingales à Temps Discrèt.* Masson, 1972.

[13] N.J. Newton. Variance reduction for simulated diffusions. *SIAM J. Appl. Math.*, 54.

[14] G.CH. Plufg. *Optimization of stochastic models.* Kluwer Academic Publisher, 1996.

[15] D. Williams. *Probability with Martingales.* Cambridge University Press, 1995.

7 Appendix

We briefly present here the Chen's projection method. For more details, one can see [7].

7.1 The hypothesis

Let $h : \mathbb{R}^m \to \mathbb{R}$ be an unknown function. We suppose that h is continuous and that $h(x^*) = 0$. Let $(X_n)_n$ be a sequence for approximating x^* and which is based on some measurements $(Y_n)_n$ of a random observation. At time $(n+1)$, the regression function h is observed at X_n with a random error ϵ_{n+1} given by

$$Y_{n+1} = h(X_n) + \epsilon_{n+1}, \quad n \geq 0. \tag{21}$$

The authors in [7] make the following hypothesis

(A) $\quad \overline{\lim} \frac{1}{n} \| \sum_{i=0}^{n-1} \epsilon_{i+1} \| = 0 \quad p.s.,$

(B) $\quad \exists \quad v : \mathbb{R}^m \to \mathbb{R},$ twice continuously differentiable such that

$$v(x^*) = 0, \quad \lim_{\|x\| \to \infty} v(x) = +\infty$$

$$\text{and} \quad v(x) > 0, \quad h(x) \cdot \nabla v(x) > 0, \quad \forall \quad x \neq x^*.$$

Remark 7.1. v is an arbitrary Lyapounov function satisfying hypothesis (B). In our case, this function was given by $v(x) = \|x - x^*\|^2$.

Remark 7.2. Condition (A) is satisfied by a large class of random vector such as ARMA processes. In addition, by Kronecker's lemma if $\sum_{i=1}^{n} \frac{1}{i} \epsilon_i$ converges $a.s.$ then condition (A) holds.

7.2 "Chen's Projection" (see [7])

To make use of their method, the authors in [7] choose $x^1 \neq x^2$ in $\mathbb{R}^m$ and fix $M > 0$ such that :

$$\max(v(x^1), v(x^2)) < \min(M, \inf(v(x); \|x\| > M)). \tag{22}$$

Afterwards, they consider an increasing sequence $(U_n)_n$ of positive numbers tending to infinity with $U_0 > M + 8$. Then they define for $n = 1, 2, \ldots$

$$X_{n+1} = \begin{cases} X_n - \frac{1}{n} Y_{n+1} & \text{if} \quad \|X_n - \frac{1}{n} Y_{n+1}\| \leq U_{\sigma(n)}, \\ x_n^* & \text{otherwise}, \end{cases} \tag{23}$$

where

$$\sigma(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\|X_k - \frac{1}{k}Y_{k+1}\| > U_{\sigma(k)}}, \quad \sigma(1) = 0,$$

$$x_n^* = \begin{cases} x^1 & \text{if } \sigma(n) \text{ is even,} \\ x^2 & \text{otherwise} \end{cases}.$$

Remark 7.3. Indeed, it is possible to find the constant M such that (22) holds, since $v(x) \to +\infty \quad$ when $\quad \|x\| \to +\infty$.
This technique of projection makes the mean field h much more stable without modifying it.

The following theorem is their main result and is very powerfull.

Theorem 6. *Under hypothesis (A) and (B), the RM algorithm defined by (23) converges a.s. to x^* and the number of truncations $\sigma(n)$ is bounded.*

Remark 7.4. This result is proved in [7]. It is important to emphasize that there is no a priori boundedness assumption imposed on X_n since the sequence $(U_n)_{n\geq 0}$ is time varying and allowed to increase to infinity.