Adaptative Monte Carlo Method, A Variance Reduction Technique

Bouhari AROUNA

CERMICS, Ecole Nationale des Ponts et Chaussées,

77455 Marne La Vallée,6 et 8 av Blaise Pascal,

France, email : arouna@cermics.enpc.fr

Abstract :

In this paper we develop an adaptative method for Monte Carlo simulations of expectations driven by high-dimensional vectors. The method uses *importance sampling* based on a change of drift. When there exist an –unknown– optimal drift, The change of drift is selected adaptatively through the Monte Carlo computation by using a suitable sequence of approximation. In the case we use a recursive non-deterministic sequence of approximation to the optimal drift, the independence of the simulated sample paths is shattered. However it is proved that we still have a Law of Large Numbers and a Central Limit Theorem for the problem. In our applications the method is used to perform variance reduction in a Monte Carlo computation of both path-dependent and path-independent options, stochastic volatility models, and interest rate derivatives.

KEY WORDS: Monte Carlo methods,variance reduction, Importance Sampling, Stochastic algorithms, CLT for martingales, Chen's projection method.

1   Introduction

Monte Carlo simulation is widely used in many areas of applied sciences. In Finance, it is frequently the only method available for the pricing and the hedging of complex path-dependent and –path-independent– options, particularly if the number of relevant asset involved is large or if additional randomness are included in the model. In all such cases, the computational demands of simulation have motivated interest in Monte Carlo methods for increased efficiency. That is to say with important reduction on variance.
In this paper we develop a Monte Carlo method which integrate an *importance sampling* technique based on an adaptative change of drift. The change of drift is selected adaptatively by using a sequence of approximation of an optimal drift. When a recursive non-deterministic sequence is used to approximate this optimal drift, the independence of the simulated sample

paths is blured. However it is proved that we still have a Law of Large Numbers and a Central Limit Theorem leading to an adaptative *"optimal"* Monte Carlo method.

In section 2 we recall without proofs the Law of Large Numbers and the Central Limit Theorem for martingales we will use to prove our main result. Section 3 deals with a general theorem which is the main result of this paper. In Section 4 we show how the result can be used to reduce variance in a Monte Carlo simulation. The last section illustrates the use of the method by some numerical applications to the pricing of options and interest rate derivatives.

2   Limit theorem for martingales

We start the work by recalling some widely known classical results on martingales: the so called law of large numbers and central limit theorem for martingales are the keys of this work.

Theorem 1. *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 a 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 assumption applies:*

(1) $\quad \dfrac{\langle\ M\ \rangle_n}{a_n} \xrightarrow[n]{\mathbb{P}} \sigma^2 \quad with \quad \sigma > 0;$

then: $\dfrac{M_n}{a_n} \xrightarrow[n]{a.s.} 0.$ *If in addition*

(2) Lindberg's condition holds; in other words, for all $\varepsilon > 0$,

$\frac{1}{a_n} \sum_{k=1}^{n} \mathbb{E}\left[|M_k - M_{k-1}|^2 1_{\{|M_k - M_{k-1}| \geq \varepsilon \sqrt{a_n}\}} / \mathcal{F}_{k-1} \right] \xrightarrow[n]{\mathbb{P}} 0,$

then: $\dfrac{M_n}{\sqrt{a_n}} \xrightarrow[n]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$

Remark 2.1. The theorem is proved in various books which give it its due importance (see [14] for example). For a deeper discussion on the subject, many references can be found in [11]. The Lindberg condition is essential to prove the last part of the theorem.

We will need also the following elementary and usefull lemma.

Lemma 1. *(Cesaro) If f is a continuous function from $\mathbb{R}^d$ to $\mathbb{R}$ and $(x_n)_{n \geq 0}$ is a sequence of real vectors which converges to x, then*

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

3 The framework

In many areas of applied sciences, it is question of computating by simulation quantities of the form

$$V = \mathbb{E}[\varphi(X)], \quad X \sim \mu(dx),$$

where $\mu(dx)$ is a given probability distribution on $\mathbb{R}^d$, $d > 1$. Very often, one need to reduce the statistical error of the Monte Carlo estimation of this expectation. There exist many classical methods which work in specific cases (see for instance [4], or [12] and the references given there). There exist also some more advanced way to reduce this error which in general are closely related to the expression of φ and the law $\mu(dx)$ of X (see among all [15], [2], [5] and [3]).

The method we propose here lies on a relatively general theorem which is the main result of this work. In order to prove this result in a general setting, let us consider a random vector Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $(Z_n)_{n \geq 0}$ be a sequence of independent vectors drawn from the law of Z.

For a fix parameter $\theta^* \in \mathbb{R}^d$ assume that there exist a sequence $(\theta_n)_{n \geq 0}$ of approximation to θ^* defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For more convenient Γ^* stands for $\{\theta_n \xrightarrow{a.s.} \theta^*\}$.

We also assume that there exist a map

$$g : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$$

such that the following conditions hold:

(A) $g(\cdot, Z) \in \mathrm{L}^4(\mathbb{P})$ and $\mathbb{E}[g(\theta_n, Z)] = \mathbb{E}[g(\theta^*, Z)]$ $\forall n \geq 0$.

(B) The map $x \longmapsto \mathbb{E}[g^p(x, Z)] \equiv s_p(x)$ is continuous at θ^*
for $p = 2, 3, 4$.

(C) $(\theta_n)_{n \geq 0}$ is adapted to the filtration of $(Z_n)_{n \geq 0}$ and
for each $n \geq 0$, $s_4(\theta_n) \leq C_n$ for some $C_n \in \mathbb{R}_+$.

Remark 3.1. Note that we do not specify anything else more about the sequence $(\theta_n)_{n \geq 0}$. It could be either deterministic or stochastic. In addition the sequence $(C_n)_{n \geq 0}$ is allowed to tend to infinity.

Remark 3.2. The last part of hypothesis (A) holds very often. For example if $Z \sim f(z)dz$, the random vector $Z + \theta$, $\theta \in \mathbb{R}^d$ has the probability density $f_\theta(z) = f(z - \theta)$. Thus for a positive measurable function φ, we can certainly write

$$\mathbb{E}[\varphi(Z)] = \mathbb{E}[g(\theta, Z)]$$

with $g(\theta, Z) = \varphi(Z + \theta)\frac{f(Z+\theta)}{f(Z)}$ and (A) is satisfied if φ is sufficiently integrable.

3

Our main result is the following

Theorem 2. *Under assumptions (A), (B) and (C),*

1. *If $\mathbb{P}(\Gamma^*) > 0$ then a.s. on Γ^* we have,*

$$S_n \equiv \frac{1}{n} \sum_{k=1}^{n} g(\theta_{k-1}, Z_k) \xrightarrow[n]{} s_1(\theta^*) = \mathbb{E}[g(\theta^*, Z)]$$

2. *If $\mathbb{P}[\Gamma^*] = 1$ then we have*

$$\sqrt{n}[S_n - s_1(\theta^*)] \xrightarrow[n]{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad with \quad \sigma^2 = \mathrm{var}[g(\theta^*, Z)].$$

3. *If we write σ_n^2 for the value of $\frac{1}{n} \sum_{k=1}^{n} g^2(\theta_{k-1}, Z_k) - S_n^2$, then*

$$\sqrt{n} \frac{[S_n - s_1(\theta^*)]}{\sigma_n} \xrightarrow[n]{\mathcal{L}} \mathcal{N}(0, 1).$$

Proof

The proof consists in the construction of a martingale for which the hypotheses of *Theorem 1.* hold.

stage 1.

We first suppose that $\mathbb{P}(\Gamma^*) > 0$ and we consider the following sequence $(M_n)_{n \geq 0}$ defined by

$$M_n = \sum_{k=1}^{n} [g(\theta_{k-1}, Z_k) - s_1(\theta_*)], \quad n \geq 1, \quad and \quad M_0 = 0.$$

We denote by $\mathcal{F}_n$ the σ-algebra generated by θ_k and Z_k for $k \leq n$. Since the sequence $(\theta_n)_{n \geq 0}$ is adapted to the natural filtration of $(Z_n)_{n \geq 0}$, it is easily seen that θ_{n-1} is $\mathcal{F}_{n-1}$-mesurable and that Z_n is independent from $\mathcal{F}_{n-1}$. Thus

$$\mathbb{E}[g(\theta_{n-1}, Z_n) / \mathcal{F}_{n-1}] = s_1(\theta_{n-1}) = s_1(\theta_*),$$

and $(M_n)_{n \geq 0}$ is a martingale with respect to $(\mathcal{F}_n)_{n \geq 0}$. By hypothesis (C) we have

$$\mathbb{E}(|\Delta M_n|^2) \leq \mathbb{E}(s_2(\theta_{n-1})) \leq \sqrt{C_n} \quad on \quad \Gamma^*,$$

so that the martingale $(M_n)_{n \geq 0}$ is squared integrable on Γ^*. Its angle bracket process is given by

$$\begin{aligned}
\langle M \rangle_n &= \sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^2 / \mathcal{F}_{k-1}] \\
&= \sum_{k=1}^{n} \left(\mathbb{E}[g^2(\theta_{k-1}, Z_k) / \mathcal{F}_{k-1}] - s_1^2(\theta_{k-1}) \right) \\
&= \sum_{k=1}^{n} \left(s_2(\theta_{k-1}) - s_1^2(\theta_{k-1}) \right).
\end{aligned}$$

4

Then using the Cesaro lemma (stated in section 2 above) we see that $a.s.$ on Γ^*

$$\frac{\langle M \rangle_n}{n} \xrightarrow[n]{} s_2(\theta^*) - s_1^2(\theta^*) = \mathrm{var}(g^2(\theta^*, Z)) \equiv \sigma^2,$$

By applying the part (1) of $Theorem\ 1.$ we have

$$a.s. \quad \text{on} \quad \Gamma^* \quad \frac{M_n}{n} \xrightarrow[n]{} 0,$$

which is equivalent to the first part of the theorem. In order to obtain the second part of the theorem, it is sufficient to prove the Lindeberg conditon for the martingale $(M_n)_{n \geq 0}$. From here to the end of the proof, we suppose that $\mathbb{P}(\Gamma^*) = 1$ and all the work below is done on the event Γ^*.
$stage\ 2.$
By few algebra one can see that

$$\mathbb{E}\left(|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})|^4 / \mathcal{F}_{k-1} \right) = \mathbb{E}\left(g^4(\theta_{k-1}, Z_k) / \mathcal{F}_{k-1} \right) - 3s_1^4(\theta_{k-1})$$

$$- 4s_1(\theta_{k-1}) \mathbb{E}\left(g^3(\theta_{k-1}, Z_k) / \mathcal{F}_{k-1} \right)$$

$$+ 6s_1^2(\theta_{k-1}) \mathbb{E}\left(g^2(\theta_{k-1}, Z_k) / \mathcal{F}_{k-1} \right).$$

Using again Cesaro's lemma we find that

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})|^4 / \mathcal{F}_{k-1} \right) \xrightarrow[n]{} L,$$

where
$$L = s_4(\theta^*) - 3s_1^4(\theta^*) - 4s_1(\theta^*)s_3(\theta^*) + 6s_1^2(\theta^*)s_2^2(\theta^*)$$

is a non–negative constant number. Now for a fix $a > 0$ and $n \geq 1$ define

$$F_n(a) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})|^2 \mathbf{1}_{\{|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})| > a\}} / \mathcal{F}_{k-1} \right).$$

We have certainly

$$F_n(a) \leq \frac{a^{-2}}{n} \sum_{k=1}^{n} \mathbb{E}\left(|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})|^4 / \mathcal{F}_{k-1} \right).$$

Then by choosing $a_n = \varepsilon \sqrt{n}$ for an arbitrary $\varepsilon > 0$ we can write

$$F_n(a_n) \leq \frac{\varepsilon}{n^2} \sum_{k=1}^{n} \mathbb{E}\left(|g(\theta_{k-1}, Z_k) - s_1(\theta_{k-1})|^4 / \mathcal{F}_{k-1} \right),$$

showing that
$$\limsup_{n \to +\infty} F_n(a_n) = 0,$$
and that the Lindeberg condition holds. Thus by applying the part 2 of *Theorem 1.* the desired result follows.

stage 3. To end the proof, we only need to show that the following two sequences

$$\frac{1}{n} \sum_{k=1}^{n} g^2(\theta_{k-1}, Z_k) \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^{n} s_2(\theta_{k-1}) \quad \text{for} \quad n \geq 1$$

converge to the same limit. Similarly to that in the *stage 1.* we define

$$\overline{M}_n = \sum_{k=1}^{n} [g^2(\theta_{k-1}, Z_k) - s_2(\theta_{k-1})], \quad n \geq 1, \, M_0 = 0.$$

The same reasoning as in *stage 1.* applies and shows that $(\overline{M}_n)_{n \geq 0}$ is a squared integrable martingale with respect to $(\mathcal{F}_n)_{n \geq 0}$. Its hook is given by

$$\langle \overline{M} \rangle_n = \sum_{k=1}^{n} \left(\mathbb{E}[g^4(\theta_{k-1}, Z_k)/\mathcal{F}_{k-1}] - s_2^2(\theta_{k-1}) \right)$$
$$= \sum_{k=1}^{n} \left(s_4(\theta_{k-1}) - s_2^2(\theta_{k-1}) \right).$$

Again by the Cesaro Lemma we easily see that

$$\frac{\langle \overline{M} \rangle_n}{n} \xrightarrow[n]{} s_4(\theta^*) - s_2^2(\theta^*) = \text{var}(g^2(\theta^*, Z)).$$

Thus we have $\frac{\overline{M}_n}{n} \xrightarrow[n]{} 0$ which equivalently can be rewrite as

$$\sigma_n^2 = \frac{1}{n} \sum_{k=1}^{n} g^2(\theta_{k-1}, Z_k) - S_n^2 \xrightarrow[n]{} s_2(\theta^*) - s_1^2(\theta^*) = \text{var}(g(\theta^*, Z),$$

and the proof is complete. ∎

Now our objective is to show how *Theorem 2.* can be used to reduce variance in Monte Carlo simulations.

4 Applications

In order to ensure hypothesis (A) of section 3, all our applications will lie in an importance sampling method.

4.1 Importance Sampling

Consider the general problem of estimating $V = \mathbb{E}[\varphi(Z)]$, for some $\varphi : \mathbb{R}^d \rightarrow [0, \infty)$, with Z a d-dimensional random vector having –multivariate– density $f(z)$. Shortly, we limit ourselves to a parametric change of law. Then

$$
\begin{aligned}
\mathbb{E}[\varphi(Z)] &= \int \varphi(z) f(z) dz \\
&= \int \varphi(z + \theta) f(z + \theta) dz \\
&= \int \varphi(z + \theta) \frac{f(z + \theta)}{f(z)} f(z) dz \\
&= \mathbb{E}\left[\varphi(Z + \theta) \frac{f(Z + \theta)}{f(Z)} \right],
\end{aligned}
$$

and among all θ the optimal one solves the problem

$$
\min_{\theta} \mathbb{E}\left[\varphi^2(Z + \theta) \frac{f^2(Z + \theta)}{f^2(Z)} \right]. \tag{1}
$$

In practice, finding the optimal θ is infeasible and even if this optimal θ can be found, it will not in general provide a zero-variance estimator (precisely because our change of law is *parametric*!). Algorithms such as gradient's algorithm, Newton's algorithm or Robbins-Monro's algorithms could be used to solve asymptotically the problem of above (see for instance [10], [11], [13] or [17]).

4.2 Finance

In a complete market the arbitrage price of an option with payoff $\psi(S_t, t \leq T)$ is given by

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

where the underlying asset S is supposed to follow the stochastic differential equation

$$
dS_t = S_t(rdt + \sigma(t, S_t)dW_t), \ S_0 = x, \tag{3}
$$

with r the risk-free, continuously compounded interest rate, $\sigma(t, y)$ the asset volatility, W a Brownian motion, and x fixed.

When an exact solution of (3) is not available we assume that an acceptable discretization of this equation has already been determined on a discrete grid of points $0 = t_0 < t_1 < \cdots < t_d = T$, 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

$$
\hat{V}_0 = \mathbb{E}[e^{-rT} \hat{\psi}(S_{t_1}, \ldots, S_{t_d})],
$$

which we rewrite as
$$\hat{V}_0 = \mathbb{E}[\varphi(Z)], \tag{4}$$
where $Z = (Z_1, \ldots, Z_d) \sim \mathcal{N}(0, I_d)$ and I_d is the identity matrix of $\mathbb{R}^d$. φ is a function we can compute by using the dicretization of S and the payoff function $\hat{\psi}$ relative to the discretized problem.

In this case (1) becomes
$$\min_{\theta \in \mathbb{R}^d} H(\theta),$$
where the function H is given by

$$H(\theta) = \mathbb{E}[\varphi^2(Z + \theta)e^{-2\theta \cdot Z - \|\theta\|^2}]. \tag{5}$$

$\|x\|$ denotes the Euclidean norm of a vector $x \in \mathbb{R}^d$ and $x \cdot y$ is the inner product of two vectors $x, y \in \mathbb{R}^d$.

It is shown in [3] that when the payoff function φ is sufficiently integrable with respect to the law of Z, problem (4) has a unique solution. It is also shown that this unique solution can asymptotically be estimated by a randomly truncated version of the Robbins Monro-algorithms (for more details about stochastic algorithms see among all [10], [11], [1] or [6]). For the convenience of the reader we briefly recall the definition of this version of the Robbins Monro algorithms from [3]. Shortly said, it is defined for $n \geq 0$, by

$$\theta_{n+1} = \begin{cases} \theta_n - \gamma_{n+1} Y_{n+1} & \text{if} \quad \|\theta_n - \gamma_{n+1} Y_{n+1}\| \leq U_{\sigma(n)}, \\ x_n^* & \text{otherwise} \end{cases} \tag{6}$$

$$\sigma(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{\|\theta_k - \gamma_{k+1} Y_{k+1}\| > U_{\sigma(k)}\}}, \quad \sigma(0) = 0, \tag{7}$$

where $(U_n)_{n \geq 0}$ is an increasing sequence of positive numbers tending to infinity and $\sigma(n)$ is the number of projections done after n iterations. x_n^* is given by

$$x_n^* = \begin{cases} x^1 & \text{if } \sigma(n) \text{ is even}, \\ x^2 & \text{if } \sigma(n) \text{ is odd}, \end{cases} \tag{8}$$

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

$$\sum_{n \geq 0} \gamma_n = +\infty \quad \text{and} \quad \sum_{n \geq 0} \gamma_n^2 < +\infty. \tag{9}$$

The sequence $(Y_n)_{n \geq 0}$ of above is defined by

$$Y_{n+1} = (\theta_n - Z_{n+1})\varphi^2(Z_{n+1})e^{-\theta_n \cdot Z_{n+1} + \frac{1}{2}\|\theta_n\|^2}, \tag{10}$$

with $(Z_n)_{n \geq 0}$ a sequence of *i.i.d.* gaussian vectors following the law of Z. The following proposition allows to apply the result of *Theorem 1*.

Proposition 4.1. *If φ satisfies $\mathbb{E}(|\varphi(Z)|^{4\delta}) < +\infty$ with some $\delta > 1$, and $g(\theta, z) \equiv \varphi(Z + \theta)e^{-\theta \cdot Z - \frac{1}{2}\|\theta\|^2}$ then hypotheses (A),(B) and (C) of section (3) hold.*

    Proof. It is a simple matter that hypothesis (A) is satisfied. By the Girsanov theorem it is rather trivial to see that

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

Now suppose that $\|\theta\| \leq K$ where K is a non negative constant. With the notation $v_p(\theta, z) = \varphi^p(z)e^{-(p-1)\theta \cdot z + \frac{p-1}{2}\|\theta\|^2}$, we have

$$|v_p(\theta, z)| \leq e^{\frac{K^2}{2}(p-1)}|\varphi^p(z)|e^{K(p-1)\|z\|}.$$

Using Hölder's inequality, it follows that

$$\int |\varphi^p(z)|e^{K(p-1)\|z\|}e^{-\frac{1}{2}\|z\|^2}dz \leq \left(\int e^{\frac{\delta K}{\delta-1}(p-1)\|z\|}e^{-\frac{1}{2}\|z\|^2}dz \right)^{1-\frac{1}{\delta}} \left(\int |\varphi^{\delta p}(z)| \right.$$
$$\left. \times e^{-\frac{1}{2}\|z\|^2}dz \right)^{\frac{1}{\delta}}. \tag{11}$$

Since $\mathbb{E}(|\varphi^{4\delta}(Z)|) < \infty$, with $\delta > 1$, the Lebesgue theorem applies and the function s_p is continuous for $p = 2, 3, 4$. Hypothesis (B) is then satisfied. By definition, we have for each $n \geq 0$, $\theta_n \leq U_n$ where $(U_n)_{n\geq 0}$ is that sequence which appears in the construction of the algorithm $(\theta_n)_{n\geq 0}$. Now if we go back to the inequality (11), we easily see that

$$s_4(\theta_n) \leq C_n, \quad \text{for each} \quad n \geq 0$$

for some constant positive number C_n, and hypothesis (C) holds. ∎

5   Numerical tests and practical considerations

We begin this section by some important remarks on the convergence of the algorithm (6-10). In equation (10), Y_{n+1} is set to satisfie to following equation
$$\mathbb{E}[Y_{n+1}/\mathcal{F}_n] = \nabla H(\theta_n)$$
with H defined in (5) and $\mathcal{F}_n = \sigma\{\theta_k, Z_k, 0 \leq k \leq n\}$. Let $\nabla^2 H(\theta^*)$ be the hessian matrix of the function H at the point θ^* (it is proved in [3] that H is twice differentiable in $\mathbb{R}^d$ and that $\nabla^2 H(\theta^*)$ is a positive definite matrix). Let L stands for the highest eigenvalue of $\nabla^2 H(\theta^*)$. It is shown in

[16] that among all sequences $(\gamma_n)_{n \geq 0}$ which verified (9), the optimal ones for the algorithm (6-10) are given by

$$\gamma_n = \frac{\alpha}{n}, \, n > 1, \quad \text{with} \quad 2\alpha L > 1, \qquad (\star).$$

Unfortunately, the computation of L is infeasible in practice as we don't know the exact value of α.

However that suggests us to make the following heuristics:

• If H is likely to take high values, then α should be small and vice versa.

• In the case of Call and Put options it can be helpfull to normalized the payoff-function φ. Indeed the function to minimize is

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

and if $\hat{\varphi}(z) = \varphi(z)/S_0$, with S_0 the spot price of the underlying, then it is easily seen that

$$\underset{\theta}{\text{argmin}} H(\theta) = \underset{\theta}{\text{argmin}} \hat{H}(\theta),$$

where

$$\hat{H}(\theta) = \mathbb{E}[\hat{\varphi}^2(Z)e^{-\theta \cdot Z + \frac{1}{2}\|\theta\|^2}].$$

The idea is thus to make use of the Robbins-Monro algorithm to compute the "argmin" of $\hat{H}$ instead of H. In this case, a good value of α found for an option with a given spot and strike (let say x_1 and K_1) prices must work for any other option with spot x_2 and strike K_2 whenever $\frac{x_1}{K_1} = \frac{x_2}{K_2}$.

5.1 Constant Volatility Model

We begin our tests with an application to the pricing of a European Basket Call option on d assets. The option payoff we consider is

$$\varphi(\hat{S}) = \left(\frac{1}{d}\sum_{i=1}^{d} S_T^{(i)} - K\right)_+,$$

where $S^{(i)}$ is the solution to the stochastic differential equation (under the risk neutral Probability)

$$dS_t^{(i)} = S_t^{(i)}\left(rdt - \sum_{j=1}^{d} \sigma_{ij}dW_j(t)\right), \quad t \geq 0, \quad (*)$$

with $\left(W_1(t), ..., W_d(t)\right)_{0 \leq t \leq T}$ a $d-$dimensional brownian motion, $(\sigma_{ij})_{0 \leq i,j \leq d}$ the volatility matrix and r the bank account.

The make use of the method is very easy. For more simplicity, we restrict attention to the case where the volatility surface is flat and the spot values

of all the underlying are the same and equal to S_0. Then the solution to the stochastic differential equation $(*)$ can be exactly simulated by setting

$$S_T^{(i)} = S_0 \exp\left[(r - d\frac{\sigma^2}{2})T + \sigma\sqrt{T}\sum_{j=1}^{d} Z_j\right], \quad i = 1,...,d$$

where σ stands for the flat volatility and $Z_1,...,Z_d$ are indepent standard normals. Numerical results in Table 4.1 confirm the effectiveness of the procedure. This table shows the variance ratios relative to standard Monte Carlo, using adaptatively the Robbins-Monro algorithm $(\theta_n)_{n\geq 0}$ defined by (6-10). Each variance reduction ratio is the variance per replication using standard Monte Carlo divided by the variance per replication using the method developped here.

To say it briefly, the algorithm introduced in this paper consists in approximating $\hat{V}_0$ –identified in (4)– using

$$\frac{1}{n}\sum_{k=1}^{n} \varphi(Z_k + \theta_k)e^{-\theta_k \cdot Z_k + \|\theta_k\|^2},$$

instead of the standard Monte Carlo estimator $\frac{1}{n}\sum_{k=1}^{n} \varphi(Z_k)$. The additional computational effort of the method is negligeable since the same simulated paths are used to perform the convergence of $(\theta_n)_{n\geq 0}$ –to the optimal θ^*– and the Monte Carlo computation. In return, we have a gain in computational effort greater than a factor of 10.

Table 4.1

Estimated Variance Reduction Ratios for the European basket call

	Parameters			Importance Sampling	
n	α	σ	$\frac{K}{S_0}$	Price	Variance Ratio
10	25	0.1	0.8	13.41	11
	25		1.0	7.42	12
	25		1.2	3.76	15
	2.5	0.2	0.8	17.86	18
	2.5		1.0	13.36	18
	2.5		1.2	10.03	20
20	2.5	0.1	0.8	15.14	13
	25		1.0	9.89	14
	25		1.2	6.33	17
	2.5	0.2	0.8	21.80	29
	2.5		1.0	18.06	30
	2.5		1.2	15.12	31

We use 100,000 Monte Carlo simulated paths for both d=10 and d=20. The option parameters are $S_0 = 50$, $r = 0.05$, and $T = 1.0$,.

In other words, to get the same accuracy with our method, it would be necessary to increase the number of Monte Carlo steps in the standard approach by a factor of 10 (in the worst of our cases). Since our method could be used systematically in a large class of situations, this gain is quiet important. It is worth pointing out that the method proposed here can be combined with more specific or standard variance reduction methods such as control variate, antithetic variables, stratification methods etc...

5.2 The Heston Model

Our next example is the Heston stochastic volatility model (1993):

$$\begin{cases} dS_t & = rS_t dt + \sqrt{v_t}S_t dW_t^1, \\ dv_t & = k(a - v_t)dt + \sigma\sqrt{v_t}dW_t^2, \end{cases}$$

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

$$\begin{cases} S_{t_{i+1}} & = S_{t_i}(1 + r\Delta t + \sqrt{v_{t_i}\Delta t}Z_i), \\ v_{t_{i+1}} & = v_{t_i} + k(a - v_{t_i})\Delta t + \sigma\sqrt{\Delta t v_{t_i}}(\rho Z_i + \sqrt{1 - \rho^2}Z_{d+i}), \end{cases}$$

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

Tableau 4.2

Estimated Variance Reduction Ratios for the European call in the Heston model

Parameters			Importance Sampling	
v_0	α	$\frac{K}{S_0}$	Price	Variance Ratio
0.04	10	0.8	23.59	25
	50	0.9	14.13	12
	100	1.0	6.99	14
	400	1.1	3.24	15
	400	1.2	1.43	18
0.09	10	0.8	23.62	20
	20	0.9	15.28	12
	40	1.0	9.16	12
	400	1.1	5.16	12
	400	1.2	2.99	17

We use 20,000 Monte Carlo simulated paths and $n = 50$ discretization steps. The option parameters are $S_0 = 100$, $r = 0.1$, and $T = 0.5$, $k = 2$, $a = 0.01$, $\sigma = 0.5$.

In our implementations we have taken the stochastic input to the model to be the single vector $(Z_1, \ldots, Z_{2d})$. In some respects, it might be more natural to think of two separate vectors, each of length d. Heston has given a closed form solution to the pricing of a European call option by the caracteristic functions technique (see [18]). But the example still useful as a numerical illustration. The last column in Table 4.2 contains the variance reduction ratios obtained in the Heston model for a european Call option. As in Table 4.1 the gains in computational effort are greater than a factor of 11. Table 4.3 shows numerical results for an Asian option in the same model. Like the previous examples the gain factors in this one still important (in the range of 10–20).

Tableau 4.3

Estimated Variance Reduction Ratios for the Asian call in the Heston model

Parameters			Importance Sampling	
v_0	α	$\frac{K}{S_0}$	Price	Variance Ratio
0.04	10	0.8	22.59	16
	20	0.9	13.70	17
	100	1.0	6.18	10
	1000	1.1	2.38	14
	4000	1.2	0.91	18
0.09	10	0.8	22.60	18
	25	0.9	14.18	11
	100	1.0	7.83	11
	400	1.1	3.99	13
	500	1.2	2.03	17

We use 20,000 Monte Carlo simulated paths and $n = 50$ discretization steps.

The option parameters are $S_0 = 100$, $r = 0.1$, and $T = 1$, $k = 2$, $a = 0.01$,

$\sigma = 0.5$.

In these particular cases, the estimation errors –with respect to the prices– contain both statistical and discretization errors. This last error is unrelated to our method which reduces only the statistical error.

5.3 The CIR Model

Following [15], our final example is the interest rate model of Cox, Ingersoll, and Ross (1985):

$$dr_t = k(a - r_t)dt + \sigma\sqrt{r_t}dW_t, \tag{12}$$

with W a standard brownian motion. It is shown in [8] that when $d \equiv 4ak/\sigma$ is an integer, the process $(r_t)_{t \geq 0}$ has the same law as $(\|X_t\|^2)_{t \geq 0}$, where X

is the $d-$dimensional process defined by

$$dX_t = -\frac{k}{2} X_t dt + \frac{\sigma}{2} dW_t,$$

and the components of X_0 are all equal to $\sqrt{r_0 d}$. The process X is known as *Orstein Uhlenbeck* process and on a discrete grid of points $t_j = jh$, $h > 0$ and $j = 0, ...,$ the ith coordinate of this process can be simulated exactly by setting

$$X_{t_{j+1}}^{(i)} = e^{-\frac{1}{2}kh} X_{t_j}^{(i)} + \frac{\sigma}{2} \sqrt{\frac{1}{k}(1 - e^{-kh})} Z_{(i-1)n+j}, \quad j = 0, 1, ..., n-1,$$

where $Z_1, ..., Z_{dn}$ are independent standard normal variables. In order to separate the examination of variance reduction from discretization bias unrelated to our method, we apply it to the CIR Model in the case when d is an integer. More precisely, we take d to be equal to 1. We consider an interest rate cap with total life T, a principal of S, and a cap rate of K. Suppose that the reset dates are $t_1 < t_2 < ... < t_n$ and define $t_{n+1} = T$. The cash flow of this cap is $S(r_{t_i} - K)_+$ at each payment date t_{i+1}, $i = 0, 1, ..., n$. Thus the total discounted payoff can be written as

$$S \sum_{i=1}^{n} e^{-h \sum_{j=0}^{i} r_{t_j}} \left(r_{t_i} - K \right)_+.$$

Tableau 4.4

Estimated Variance Reduction Ratios for Cap Prices in the CIR Model

Parameters			Importance Sampling	
T	α	K	Price	Variance Ratio
0.25	0.001	0.044	30.21	22
	0.001	0.054	15.31	18
	0.1	0.064	3.56	11
	1.0	0.074	0.40	20
	200	0.084	0.03	84
0.50	0.001	0.044	28.63	19
	0.001	0.054	14.98	16
	0.01	0.064	4.62	11
	0.1	0.074	1.02	16
	20	0.084	0.19	39

We use 50,000 Monte Carlo simulated paths and $n = 16$ reset dates. The cap parameters are $r_0 = 0.064$, $k = 0.05$, and $d = 1$, $k = 2$, $\sigma = 0.08$. Prices are for face value of 100.

As noted in [15] this formulation is a little bit nonstandard in that it blurs the distinction between discrete and continuous compounding (see [9] or [7] for some details). Nevertheless the example still illustrative.

In Table 4.4, the variance is reduced by factors in the range of 10–90. Contrary to the other cases, the values of α in Table 4.4 are all most smaller than 1 –except in two cases–. This is easily explained by the fact that all the prices in this table are for face value $S = 100$, so that condition $(\star)$ of above holds for relatively small α. Due to the heuristics made at the begining of this section, the same pricing problem with prices for face value $S = 1$ will give the same variance reduction ratios if each value of the parameter α in Table 4.4 is replaced with $\alpha \times 10^4$.

6 Conclusion

The framework we have developped in this paper is relative to Theorem 2 which itself still of interest. The use of this theorem is based on Conditions (A), (B), and (C). These conditions are fullfilled under some integrability assumptions –which can be weaken–, when an importance sampling method is used. However, another type of transformation could lead to an analogue result. We use this theorem in order to reduce variance in Monte Carlo estimations by combining importance sampling technique and Robbins-Monro's algorithm. The method is applicable in a very general setting and works for widely used derivatives. It doesn't cost nothing for a negligeable gain in computation. It could be used even if another variance reduction methods are available. In this case the gain could be very large.

References

[1] A. BENVENISTE, M. METIVIER, and P. PRIOURET. *Algorithmes adaptatifs et approximations stochastiques*. Masson, 1987.

[2] A.G.Z. KEMNA and A.C.F. VORST. A pricing method based on average asset values. *J. Banking Finance*.

[3] B. AROUNA. Robbins-monro algorithms and variance reduction in finance. *working paper (submitted)*.

[4] B. LAPEYRE, A. SULEM, and D. TALAY. *Understanding Numerical Analysis for Financial Models*. "outcoming", France, 2002.

[5] B. LAPEYRE and E. TEMAM. Competitive monte carlo method for the princing of asian options. *J. Computational Finance*, 5:39–57, 2001.

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

[7] J.C. HULL. *Options, Futures, and Other Derivatives*. Prentice Hall, Fifth Edition, 2003.

[8] L.C.G. ROGERS. Which model for the term structure of interest rates should one use? *Mathematical Finance*, IMA 65:93–116, 1995.

[9] M. MUSIELA and M. RUTHKOWSKI. *Martingale Method in Financial Modelling*. Springer, 1997.

[10] D. MARIE. *Algorithmes stochastiques*. Springer-Verlag, France, 1996.

[11] D. MARIE. *Random Iterative Models*. Springer-Verlag, France, 1997.

[12] N. MADRAS. *Lecture on Monte Carlo methods*. Fields Institute Monograph, AMS, 2002.

[13] N. MARTOLI and P. DEL MORAL. *Simulation et algorithmes stochastiques*. Cépaduès-éditions, 2002.

[14] J. NEVEU. *Discrete Parameter Martingales*. North Holland, 1975.

[15] P. GLASSERMAN, P. HEIDELBERGER, and P. SHAHABUDDIN. Asymptotic optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9(2):117–152, April 1999.

[16] M. PELLETIER. *Sur le comportement asymptotique des algorithmes stochastiques*. PhD Thesis, Université Paris-Sud, 1996.

[17] G.CH. PLUFG. *Optimization of stochastic models*. Kluwer Academic Publisher, 1996.

[18] S. L. HESTON. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.