

Variance reduction and Robbins-Monro algorithms

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

March, 20 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. Then, we illustrate the method by applications to options pricing.

1 Introduction

Premia 14

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 [10]. 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 [8] 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 price in advance. This approach is therefore based on using approximations of the option 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 [10]). 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. 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 also [3]).

2 Mathematical Context

2.1 Financial background

For more simplicity we use the Black and Scholes model to describe the price of an asset. However, our method can be used in a more general setting as pathdependent options or in stochastic volatility model. The price of the risky asset S_t is supposed to follow a stochastic diffusion equation under the neutral risk probability :

$$dS_t = S_t(rdt + \sigma dW_t), \quad S_0 = x,$$

which solution is given by

$$S_t = x \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right).$$

S_0 is the price of the asset at time 0. Obviously we can simulate this asset at dates $0 < T_1 < \dots < T_m = T$ using the exact representation

$$S_{T_i} = S_{T_{i-1}} \exp\left(\left(r - \frac{\sigma^2}{2}\right)(T_i - T_{i-1}) + \sigma \sqrt{T_i - T_{i-1}} Z_i\right), \quad (1)$$

where Z_1, \dots, Z_m are independent gaussian random variables with mean 0 and variance 1. We restrict attention to simulations driven by a sequence of independent normal variables, since we can recover this case when the normal variables are correlated through a linear transformation. If we want

to price an option which payoff is given by $f(S_{T_1}, \dots, S_{T_m})$ we thus have to evaluate

$$V_0 = \mathbb{E}[e^{-rT} f(S_{T_1}, \dots, S_{T_m})]. \quad (2)$$

Using (1), V_0 can be rewritted as

$$V_0 = \mathbb{E}[G(Z_1, \dots, Z_m)] , \quad (3)$$

where G is a function which can be computed and $Z = (Z_1, \dots, Z_m) \sim \mathcal{N}(0, I_m)$.

In what follows, the objective is to evaluate (3) using an importance sampling procedure.

2.2 Importance sampling

We change the law of $Z = (Z_1, \dots, Z_m)$ adding a drift vector $\mu = (\mu_1, \dots, \mu_m)$. An elementary version of Girsanov theorem applied to 3 leads to the following representation of V_0 :

$$V_0 = \mathbb{E}(\alpha(\mu)) , \quad (4)$$

with

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

where $\|x\|$ denotes the Euclidean norm of a vector $x \in \mathbb{R}^m$. The authors in [10] 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 of the fixed-point problem : $\nabla \log G(\mu) = \mu$.

In this paper, we use a RM algorithm which enables us to assess the “optimal sampling direction” μ^* that minimizes the variance of $\alpha(\mu)$, $\mu \in \mathbb{R}^m$ or equivalently :

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

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 :*

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

Proof Using a change of variables and the Girsanov theorem, we obtain

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

Suppose that $\|\mu\| \leq K$ where K is a non negative constant. With the notation

$g(\mu, z) = (\mu - z)G^2(z)e^{-\mu \cdot z + \frac{1}{2} \|\mu\|^2}$, we have

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

By Cauchy-Schwarz inequality, we can write

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

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

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

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

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

where $\text{Hess} H(\cdot)$ 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

$$\forall u \in \mathbb{R}^m - \{0\}, u^T \text{Hess} H(\mu) u = \mathbb{E} \left[\left(\|u\|^2 + (u \cdot (\mu - Z))^2 \right) G^2(Z) e^{-\mu \cdot Z + \frac{1}{2} \|\mu\|^2} \right] > 0.$$

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

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

Therefore, if $\mathbb{P}(G(Z) > 0) \neq 0$, then $\lim_{\|\mu\| \rightarrow +\infty} H(\mu) = +\infty$.

As a consequence of the proposition above, μ^* minimizing H is the unique solution of

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

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.

3.1 General framework

The RM algorithms have the form

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

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

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

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 $X_k, Y_k, k \leq n$. Clearly we can write

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

The following theorem is proved in [6] or [5].

Theorem 1. *Under the following hypothesis*

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

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

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

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 [9]. Some papers are devoted to the convergence properties of these algorithms see e.g. [2] and [4].

3.2 Application to variance reduction

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

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

where Z is drawn 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 for Y_{n+1} :

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

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

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

Therefore hypothesis (H_1) of the theorem above is satisfied. On the contrary, hypothesis (H_3) can not be satisfied. Obviously this fact is due to the exponential form of Y_{n+1} (see (16)). 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 [4] (see also [3]) 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 $n \geq 0$,

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

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

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

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

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

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

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. We propose in the last section an empirical way that to choose efficiently these coefficients.

Remark 3.2. The constant M above has no significant effect on the numerical convergence of the algorithm. In our numerical tests M values are in the interval $[10, 100]$ 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 $v : \mathbb{R}^m \rightarrow \mathbb{R}$, such that :*

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

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

2) *Let G be a function verifying $|G(z)| \leq b(1 + e^{c \cdot z})$ with $b > 0$ and $c \in \mathbb{R}^m$, then we can choose the sequence U_n such that*

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

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 $\forall u \in \mathbb{R}^m$, $H(\mu^*) < H(u)$, 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

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

It follows that

$$\begin{aligned} s^2(x) &\leq C^t e^{\|x\|^2} \left(\|x\|^2 \mathbb{E}(e^{(4c-2x) \cdot Z}) + \mathbb{E}(\|Z\|^2 e^{(4c-2x) \cdot Z}) \right) \\ &= C^t e^{\|x\|^2} \left(\|x\|^2 e^{2\|2c-x\|^2} + (m + 2\|2c-x\|^2) e^{2\|2c-x\|^2} \right) \\ &\leq C^t (1 + \|x\|^2) e^{5\|x\|^2}. \end{aligned}$$

Using equations (17-20), we get

$$\|X_n\| \leq \max(U_{\sigma(n)}, \|x_n^*\|) \leq U_n, \quad \text{for } n \text{ sufficiently large.}$$

Finally

$$s^2(X_n) \leq C^t U_n^2 e^{5U_n^2}.$$

We conclude the proof by choosing the sequence U_n such that

$$\sum_n \gamma_n^2 U_n^2 e^{5U_n^2} < +\infty.$$

Remark 3.3. The sequence U_n must increase sufficiently slow to cancel the explosion behaviour of the algorithm without modifying the mean field h . It's choice is not difficult. For example, the sequence $U_n = \sqrt{\frac{1}{6} \ln n} + U_0$, $n \geq 1$ is suitable.

Theorem 2. *In the framework of Lemma 1, the algorithm X_n defined by (17) 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 down $\epsilon_{n+1} = h(X_n) - Y_{n+1}$, $n \geq 0$ and define $M_n = \sum_{i=0}^{n-1} \gamma_{i+1} \epsilon_{i+1}$ for $n \geq 1$. The sequence $(M_n)_{n \geq 1}$ is a \mathcal{F}_n -martingale and its brackets process $\langle M \rangle$ is given by

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

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

$$\begin{aligned} \lim_{n \rightarrow +\infty} \langle M \rangle_n &\leq \sum_{n=0}^{+\infty} \gamma_{n+1}^2 \mathbb{E}[\|Y_{n+1}\|^2 / \mathcal{F}_n] \quad a.s. \\ &\leq C \sum_n \gamma_n^2 U_n^2 e^{5U_n^2} \\ &< +\infty, \end{aligned}$$

where $C > 0$. Therefore the martingale M_n converges *a.s.* and in \mathcal{L}^2 (see [7] or [1]). The Kronecker's lemma (see for example [11] p.117) implies that $\lim_{n \rightarrow +\infty} \gamma_{n+1} \|\sum_{i=0}^{n-1} \epsilon_{i+1}\| = 0$ *a.s.*. Chen, Guo and Gao proved in [3] that assuming the first part of **Lemma 1** holds, one just need the additional assumption $\lim_{n \rightarrow +\infty} \gamma_{n+1} \|\sum_{i=0}^{n-1} \epsilon_{i+1}\| = 0$ *a.s.* in order to obtain the convergence of the algorithm. **Theorem 2** is then a consequence of **Theorem 3** (see Appendix or [3]).

Remark 3.4. There are many financial cases where the payoff function G satisfies the condition $|G(z)| \leq b(1 + e^{c \cdot z})$ of the theorem such as digital, corridor, forward floor, payer swaption, etc,...

4 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 [5] or [9]) 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”. We use $\beta = 1$ for all the numerical tests. 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.

“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 more accurate than those obtained by a classical Monte Carlo method.

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 25% of the CPU time spent in the classical Monte Carlo computation. In fact we use at most 20% gaussian paths in addition to those simulate 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.

Table 4.3 shows some results about the pricing of a european basket put on 10 or 20 assets. The results obtained are very 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.

Table 4.1
Estimated Variance Reduction Ratio for the European Put

Parameters				Importance sampling			
α	β	σ	strike	RMPrice	BSPrice	CPrice	StdRatio
5.	1.	0.3	30	0.134	0.134	0.125	6.2
0.1			40	1.281	1.280	1.272	3.3
0.01			50	4.684	4.677	4.695	2.5
0.001			60	10.541	10.526	10.575	2.2
0.001			80	26.789	26.771	26.846	2.1
100.		0.1	40	0.0042	0.0042	0.0039	18.7
1.			50	0.965	0.964	0.962	3.1
0.1			60	7.310	7.305	7.327	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

Parameters				Importance sampling			
α	β	σ	strike	RMPrice	BSPrice	CPrice	StdRatio
0.01	1.	0.3	30	21.630	21.598	21.520	4.1
0.05			40	13.249	13.231	13.155	3.3
0.1			50	7.118	7.116	7.006	3.3
0.5			60	3.454	3.452	3.434	3.9
0.1			80	0.673	0.673	0.680	6.8
0.0006		0.1	30	21.466	21.463	21.520	10.6
0.001			40	11.965	11.955	11.934	4.4
0.01			50	3.405	3.402	3.379	2.8
0.07			60	0.231	0.231	0.232	5.6
5.			70	0.0035	0.0035	0.0042	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 $T = 1.0$.

Tableau 4.3
Estimated Variance Reduction Ratio for the European Basket Put

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
	0.02			70	18.41	2.1
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
	0.001			70	20.63	2.0
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
	0.001			70	24.21	1.9

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%.

We use respectively 900,000 and 100,000 simulation paths for Monte Carlo computation and Robbins Monro algorithm in this example. In this particular case only 10% additional simulation effort leads to a variance reduction with a factor of at least 4.

Table 4.4 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 for a put.

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 [10] 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.

Tableau 4.4
Estimated Variance Reduction Ratio for the Asian Put

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 $T = 1.0$.

Through our simulation results we take $c = 2$, $\nu = 0$, $r = 0.05$, $S_0 = 50$, $T = 1$, $\rho = 0.5$ et $\sqrt{\sigma_0} = 0.1$. 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 values of α . The payoff we consider is again that of a put option on the arithmetic mean

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

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. Again, we can see through these examples that the confidence interval length reduction is greater than a factor of 2.

5 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. In addition, it is 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 in function of 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 can naturally be improved. For example, one could need more stability of the algorithm relatively to the steps sequence parameters.

6 Appendix

We briefly present here the Chen's projection method. For more details, one can see [3].

6.1 The hypothesis

Let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ 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

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

The authors in [3] make the following hypothesis

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

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

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

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

Remark 6.1. v is an arbitrary Lyapounov function satisfying hypothesis (B). In our case, this function was given by $v(x) = \|x - x^*\|^2$.

Remark 6.2. 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.

6.2 “Chen's Projection” (see [3])

To make use of their method, the authors in [3] choose $x^1 \neq x^2$ in \mathbb{R}^m and fix $M > 0$ such that :

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

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

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

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 6.3. Indeed, it is possible to find the constant M such that (22) holds, since $v(x) \rightarrow +\infty$ when $\|x\| \rightarrow +\infty$. 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 3. *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 6.4. This result is proved in [3]. 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.

References

- [1] D. D.-Castelle and D. Marie. *Probabilit  s et Statistiques (tome 2.)*. Masson, France, 1983. 9
- [2] D.S. Clark . Necessary and sufficient conditions for the robbins-monro method. *Stochastic Processes and their Applications*, 17:359–367, 1984. 5
- [3] 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. 2, 6, 9, 15, 16
- [4] H.F. Chen and Y.M. Zhu. Stochastic approximation procedure with randomly varying truncations. *Scientia Sinica (series A)*, 29(9):914–926. 5, 6
- [5] D. Marie. *Algorithmes stochastiques*. Springer-Verlag, France, 1996. 5, 9

- [6] D. Marie. *Random Iterative Models*. Springer-Verlag, France, 1997. 5
- [7] J. Neveu. *Martingales À Temps Discr  t*. Masson, 1972. 9
- [8] N.J.NEWTON. Variance reduction for simulated diffusions. *SIAM J. Appl. Math.*, 54(6):1780–1805, 1994. 1
- [9] G.CH. Plu  g. *Optimization of stochastic models*. Kluwer Academic Publisher, 1996. 5, 9
- [10] P.GLASSERMAN P.HEIDELBERGER P.SHAHABUDDIN. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 2, April:117–152, 1999. 1, 2, 3, 13
- [11] D. Williams. *Probability with Martingales*. Cambridge University Press, 1995. 9