Monte Carlo Methods and stochastic approximations

Bouhari Arouna

Thèse de doctorat de l'École Nationale des Ponts et Chaussées

Spécialité : **PROBABILITÉS APPLIQUÉES**

Présentée par Bouhari AROUNA

Pour obtenir le grade de Docteur de l'École Nationale des Ponts et Chaussées

Sujet de la thèse:

Algorithmes Stochastiques et Méthodes de Monte Carlo

Soutenue le 15 Décembre 2004 devant le jury composé de :

Mr Gilles PAGES (Président, Professeur Université Paris VI)

Mr Bernard LAPEYRE (directeur de thèse, Professeur ENPC)

Mr Denis TALAY (rapporteur, Directeur de Recherche INRIA)

Mr Jean LACROIX (rapporteur, Professeur Université Paris VI)

Mr Damien LAMBERTON (examinateur, Professeur Université MLV)

Mr Stéphane TYC (examinateur, Head of Research at BNPPARIBAS)

A ma douce Emilie,
A ma très grande famille,

A tous les enfants de la Terre.

Remerciements :

C'est une vieille tradition que de remercier au début d'un tel travail tous ceux qui, de loin ou de près, directement ou indirectement, ont contribué à le rendre possible. c'est avec mon enthousiasme le plus vif et le plus sincère que je voudrais rendre mérite à tous ceux qui à leur manière m'ont aidé à mener à bien cette thèse.
Je désire exprimer ma profonde gratitude:

Au Professeur et ami Bernard LAPEYRE pour avoir accepté de me superviser durant ces trois dernières années, pour sa disponibilité et sa générosité exceptionnelles et pour son soutien constant pendant la rédaction de cette thèse. Au cours de son exercise à la direction du cermics, il a su créer une ambiance formidable et génératrice d'inspirations.

Au Professeur Gilles PAGÈS qui m'a honoré en acceptant d'être Président du jury de ma thèse et plus particulièrement pour son cours intitulé "Martingales et Algorithmes Stochastiques" du D.E.A. "Analyse et Systèmes Aléatoires" de l'Université de Marne-La-Vallée qui a constituer ma toute première expérience avec cette notion mathématique qui prends de plus en plus d'importance dans les applications.

Aux Professeurs Denis TALAY et Jean LACROIX pour avoir accepté la charge d'être Rapporteurs de ma thèse. Merci à Denis TALAY pour son intérêt pour mon travail de thèse. La semaine que j'ai passée à son invitation, dans son équipe de Sophia Antipolis a été, pour moi, extrêmement enrichissante. Merci à Jean LACROIX pour avoir accepté, sans hésitation cette tâche, alors même que nos chemins ne s'étaient jamais croisés auparavant.

Au Professeur Damien LAMBERTON et au Docteur Stéphane TYC qui m'ont honoré en acceptant d'être Examinateurs dans ce jury et d'avoir été présents à ma soutenance de thèse. Chacun d'eux mérite un remerciement plus particulier :

Damien pour toutes ses remarques, ses suggestions et ses conseils amicaux, cela m'a beaucoup encouragé; je dis merci à l'homme, mais aussi à l'enseignant qu'il a été pour moi.
Stéphane de m'avoir accepté dans son équipe de recherche en finance quantitative à la BNPPARIBAS alors même que je n'avais pas encore soutenu ma thèse: merci.

Je tiens à remercier tous les membres de l'équipe Probabilités Appliquées du Cermics, notamment Jean-François DELMAS, Benjamin JOURDAIN, Jérome LELONG, Nicola MORENI, Aurélien ALFONSI, Christophe CHORRO, Julien GUYON, Ralph LAVIOLETTE et Antonino ZANETTE, pour leur gentillesse et les discussions stimulantes que nous avons eues durant ces trois dernières années.

Je tiens également à remercier tous les autres membres du cermics: Alexandre ERN, Linda EL ALAOUI, Laetitia ANDRIEU, Kengy BARTY, Jean Philippe CHANCELIER,Guy COHEN, Eric CANCES, Yousra GATI, Claude LE BRIS, Frédéric LEGOLL, Tony LELIEVRE, François LODIER ainsi que tous les autres dont je n'ai pas cité le nom ici pour leur ouverture d'esprit et leur gentillesse. Nos parties de pétanque entre "thésards" me manquerons beaucoup.

Ce ne serait pas tout à fait correct de ma part de finir sans adresser mes plus vifs remerciements à Sylvie BERTE et Khadija ELOUALI, les deux secrétaires du Cermics pour m'avoir supporter pendant ces trois années. Leur tâche est immense et leur présence est rassurante pour nous, même si nous ne le leur disons pas souvent. Avec du recul, je vous dis chapeau mesdames.

Je dois un grand merci à tous mes amis, et en particulier à Latyr MBODJ, pour avoir accepté de relire certaines parties de cette thèse, durant ses week ends et à sa compagne Josiane ZOUNGRANA d'avoir supporté, avec souplesse, cette situation où il lui consacre forcément un peu moins de temps.

Je ne pourrais jamais oublier le soutien et l'aide des membres de ma nombreuse et merveilleuse famille. Et que dire à ma chère et tendre moitié, Emilie NABEDE, elle qui m'a toujours supporté contre vents et marrées, sinon un sobre et simple *merci infiniment*.

Table des Matières

Introduction

Introduction

Le but de ce travail de thèse est de proposer des techniques de réduction de variance, facilement implémentables, pour les méthodes de Monte Carlo. Notre approche se fait en deux étapes:

- La paramétrisation de la loi de simulation à travers un changement de tendance (drift),
- L'utilisation des Algorithmes Stochastiques pour estimer le paramètre de drift optimal.

Nous avons d'abord développé une implémentation séquentielle de la méthode, suivie ensuite d'une implémentation adaptative. Cette dernière technique sort du cadre habituel des applications des méthodes de Monte Carlo. Nous avons ainsi été amené à mettre en place et à démontrer des résultats théoriques de convergence qui définissent le cadre précis de l'utilisation de cette méthode de réduction de variance adaptative. Voici schématiquement comment se présente les deux approches séquentielle et adaptative. Lorsque l'on cherche à calculer la quantité

$$\mathbb{E}[\varphi(X)],$$

une simulation Monte Carlo s'avère souvent être la seule méthode ayant un facteur "gain/coût (temps de calcul)" raisonnable. C'est le cas, par exemple, lorsque X est un vecteur aléatoire de très grande dimension (> 10). Si l'on dispose d'une représentation paramétrique de l'espérance sous la forme:

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[g(\theta, X)], \quad \theta \in \mathbb{R}^d,$$

il est alors naturel d'essayer d'identifier un paramètre optimal θ^* qui minimise l'erreur statistique ou la variance

$$\mathbb{E}[g^2(\theta, X)] - \mathbb{E}[\varphi(X)]^2$$

de l'estimation. Les Algorithmes (ou Approximations) Stochastiques de type Robbins–Monro sont bien adaptés pour ce type d'optimisation stochastique. Ils permettent d'obtenir une estimation asymptôtique du paramètre θ^*. Il suffit ensuite d'utiliser la Loi Forte des Grands Nombres pour estimer l'espérance recherchée par

$$\frac{1}{N}\sum_{n=1}^{N} g(\theta^*, X_n),$$

où $(X_n)_{n\leq N}$ est une suite de vecteurs aléatoires indépendants tous de même loi que X et N un entier suffisamment grand. Nous avons montré que si l'on sait construire une

suite $(\theta_n)_{n\geq 0}$ qui converge presque sûrement vers θ^*, l'on peut utiliser l'approximation non standard suivante

$$\mathbb{E}[\varphi(X)] \simeq \frac{1}{N}\sum_{n=1}^{N} g(\theta_{n-1}, X_n).$$

Nous obtenons également un Théorème de la Limite Centrale et un résultat d'estimation de la variance empirique pour cette approximation et développons ainsi une méthodologie Monte Carlo non standard[1]. Notons qu'il est possible de démontrer un résultat de type Berry Essen pour cette méthode de Monte Carlo non standard.

Présentation du Résultat Adaptatif

On veut calculer à l'aide d'une méthode de Monte Carlo la quantité $\mathbb{E}[\varphi(X)]$, pour une fonction (éventuellement vectorielle) $\varphi : \mathbb{R}^d \to \mathbb{R}$ suffisamment intégrable et un vecteur aléatoire $d-$dimensionnel X defini sur un espace probabilisé donné $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose qu'il existe une transformation

$$g : \mathbb{R}^q \times \mathbb{R}^d \longrightarrow \mathbb{R}, \qquad q \geq 1$$

suffisamment intégrable telle que:

(A) $\quad \mathbb{E}[g(\theta, X)] = \mathbb{E}[\varphi(X)], \quad \forall\ \theta\ \in \mathbb{R}^d.$

Soit θ^* un point donné de $\mathbb{R}^q$ et $(\theta_n)_{n\geq 0}$ une suite de vecteurs aléatoires définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et convergeant presque sûrement vers θ^*. On montre sous des hypothèses assez faibles que si $(X_n)_{n\geq 0}$ est une suite de vecteurs aléatoires indépendants identiquement distribués selon la loi de X alors:

le vecteur aléatoire

$$S_N \triangleq \frac{1}{N}\sum_{n=1}^{N} g(\theta_{n-1}, X_n)$$

converge presque sûrement, quand N tend vers l'infinie, vers l'espérance recherchée

$$\mathbb{E}[g(\theta^*, X)] = \mathbb{E}[\varphi(X)].$$

Sous les mêmes hypothèses, on montre qu'un théorème de la limite centrale existe:

[1] le domaine d'application principal des résultats théoriques de ce travail est la finance numérique. Toutefois, il nous semble que les techniques développées sont suffisamment génériques pour être adaptées en vu de leur utilisation dans d'autres domaines comme par exemple la physique.

$$\sqrt{N}[S_N - \mathbb{E}[\varphi(X)]] \underset{N\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, \Sigma^2),$$

avec Σ^2 une matrice de variance-covariance qu'on peut expliciter en fonction de θ^*. Sous des hypothèses légèrement renforcées, on montre l'existence d'intervalles de confiance empiriques pour ce théorème de la limite centrale.

Ces résultats, qui sont exposés dans le Chapitre 3, sont à la base de notre méthodologie d'estimation Monte Carlo dans laquelle la variance empirique de l'estimation est réduite dynamiquement au cours des itérations Monte Carlo.

Mise en œuvre de la Méthode

L'exemple du cas Gaussien

En finance l'implémentation du prix d'une option peut le plus ouvent se ramener au calcul de $\mathbb{E}[\varphi(X)]$ où φ est la fonction de payoff de l'option et $X = (X_1, \ldots, X_d) \sim \mathcal{N}(0, I_d)$ est un vecteur Gaussien standard. I_d désigne la matrice identité de $\mathbb{R}^d$. Un simple changement de variable entraine l'égalité

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[g(\theta, X)], \quad \text{avec} \quad g(\theta, x) = \varphi(x+\theta)e^{-\theta\cdot x - \frac{1}{2}\|\theta\|^2}, \quad \theta \in \mathbb{R}^d,$$

où $\|x\|$ représente la norme Euclidienne d'un vecteur x appartenant à $\mathbb{R}^d$ et $x\cdot y$ désigne le produit scalaire de deux vecteurs $x, y \in \mathbb{R}^d$. On note $V(\theta) = \mathbb{E}[g^2(\theta, X)] - \mathbb{E}[\varphi(X)]^2$. On montre sans difficulté particulère que si

la fonction de payoff $\varphi(X)$ est non nulle et suffisamment intégrable alors la fonction $V(\theta)$ est différentiable et possède un unique minimum θ^*,

unique solution du problème de recherche du zéro du gradient de V

$$\nabla V(\theta) = 0.$$

A cause de la forme "exponentielle" de $\nabla V(\cdot)$, l'algorithme de Robbins–Monro classique ne permet pas de résoudre directement cette équation. Pour cela nous utilisons une méthode de projection aléatoire de l'algorithme de Robbins-Monro dûe à Chen et Zhu (1986). En exploitant cette méthode, on a montré que sous la seule condition d'intégrabilité de la fonction de payoff:

$$\mathbb{E}(|\varphi(X)|^{4\delta}) < +\infty \quad \text{pour un certain} \quad \delta > 1,$$

on peut choisir une version $(\theta_n)_{n\geq 0}$ de l'algorithme de Robbins-Monro convenablement tronqué par la méthode de Chen et Zhu qui converge presque sûrement vers l'unique racine θ^* du problème $\nabla V(\theta) = 0$, $\theta \in \mathbb{R}^d$.

L'algorithme de Robbins–Monro projeté par cette méthode de Chen est présenté dans le chapitre 1 et le résultat ci-dessus est exposé dans le Chapitre 2.

Dans ce Chapitre notre motivation a ensuite été d'illustrer sur des exemples tirés de la finance l'intérêt numérique de l'approximation

$$\mathbb{E}[\varphi(X)] \simeq \frac{1}{N}\sum_{n=1}^{N} g(\theta_*, X_n),$$

en comparant sa variance estimée et son coût (en temps de calcul) à celle d'une méthode de Monte Carlo standard et à celle de la méthode des fonctions d'importance de Glasserman, Heidelberger et Shahabuddin (1999). Ce Chapitre fait l'objet d'un article paru dans la revue **"The Journal of Computational Finance"**.

Dans le Chapitre 3 nous étudions l'approximation Monte Carlo non standard

$$\mathbb{E}[\varphi(X)] \simeq \frac{1}{N}\sum_{n=1}^{N} g(\theta_{n-1}, X_n).$$

Dans la première partie de ce chapitre, nous utilisons les algorithmes de Robbins–Monro à pas décroissant pour développer une méthode de Monte Carlo adaptative. Dans cette méthode, la variance est réduite au fur et à mésure des itérations Monte Carlo pour finalement donner un estimateur de variance minimale en un certain sens. Cette partie du Chapitre 3 a constitué un article publié dans la revue **"Monte Carlo Methods and Applications"**.
Dans la deuxième partie de ce chapitre, nous reprenons les méthodes développées au chapitre 2, en les adaptant aux algorithmes stochastiques à pas constants $\gamma > 0$ (mais petits). La première remarque à faire est que ces algorithmes ne convergent pas *presque sûrement* vers θ^* mais convergent en loi vers un vecteur aléatoire qu'on peut écrire sous la forme

$$G = \theta^* + \sqrt{\gamma}\, g,$$

avec g un vecteur gaussien centré. Ainsi la loi du vecteur G se concentre autour de θ^* lorsque le pas γ devient de plus en plus petit. C'est cette observation qui justifie

(moralement) le fait qu'on puisse utiliser une réalisation de la loi stationnaire des algorithmes stochastiques à pas constants à la place de θ^* dans les procédures de réduction de variance.

Dans le Chapitre 4, nous considérons le problème de réduction de variance sous un angle différent: celui de la minimisation de l'entropie relative entre une certaine probabilité optimale et une famille paramétrique de probabilités obtenue par le Théorème de Girsanov. Nous montrons que ce point de vue conduit (en univers discret) à la minimisation de la fonction suivante:

$$V(\theta) = \mathbb{E}\left[\varphi(X)\left(\frac{1}{2}\|\theta\|^2 - \theta \cdot X\right)\right], \quad \theta \in \mathbb{R}^d.$$

Il est clair que cela revient au même de chercher le zéro $\hat{\theta}$, de la fonction

$$\nabla V(\theta) = \mathbb{E}\left[\varphi(X)\left(\theta - X\right)\right], \quad \theta \in \mathbb{R}^d.$$

A cause de la disparition du terme $e^{-\theta \cdot X + \frac{1}{2}\|\theta\|^2}$ Ce dernier problème est de loin plus facile à résoudre que la recherche de zéro de la fonction

$$\nabla V(\theta) = \mathbb{E}[(\theta - X)\varphi^2(X)e^{-\theta \cdot X + \frac{1}{2}\|\theta\|^2}].$$

Manifestement on a $\hat{\theta} = \frac{\mathbb{E}[\varphi(X)]}{\mathbb{E}[X\varphi(X)]}$ et l'on peut ainsi simuler $\hat{\theta}$ directement.
Mais cette manière de procéder ne permet pas de mettre en place une méthode de réduction de variance adaptative. Nous avons donc utilisé de nouveau des algorithmes stochastiques pour estimer $\hat{\theta}$. Et nous avons surtout remarqué que le paramètre $\hat{\theta}$ permettait de construire un estimateur convergeant et consistant du delta de l'option dont le payoff est $\varphi(\cdot)$. Cette possibilité qu'a l'algorithme de fournir une estimation de la couverture delta[2] de l'option est particulièrement intéressante.

[2]le delta d'une option est la dérivé première de son prix par rapport au titre sur lequel l'option porte. C'est cette quantité qui dit "de combien le prix de l'option varie lorsque le sous-jacent bouge".

Chapitre 1

Rappels Mathématiques

1.1 Contenu du chapitre

Dans ce chapitre nous présentons certains résultats et outils théoriques bien connus qui nous seront utiles tout au long de cette thèse. Nous commençons, dans les sections 1 et 2, avec une brève présentation des grandes lignes de la méthode de Monte Carlo et les résultats de convergence classique qui définissent son cadre d'utilisation. Dans la section 3 nous rappelons les théorèmes limites pour martingales que nous utiliserons dans les chapitres suivants. Par la suite nous introduisons les algorithmes stochastiques sous leur forme générale suivi par des résultats de convergence que nous commentons. Pour finir nous présentons la technique de projection fixe et la technique de projection aléatoire de Chen qui sont utilisées dans les applications pratiques des algorithmes stochastiques.

1.2 Présentation de la méthode de Monte carlo

Le recours aux méthodes de Monte Carlo est très courant chaque fois que l'on est confronté au calcul de certaines quantités moyennes. Après avoir mis ces dernières sous forme d'espérances, il reste à calculer des quantités de la forme $\mathbb{E}[\varphi(X)]$, où X est un vecteur aléatoire et φ une fonction définie sur $\mathbb{R}^d$. Pour pouvoir faire ce calcul, il convient de savoir *simuler* une suite $(X_n,\ n \geq 1)$ de variables aléatoires indépendantes selon la loi de X.
Sur le plan informatique, on ramène la simulation d'une loi arbitraire à celle d'une suite de variables aléatoires indépendantes uniformément distribuées sur l'intervalle $(0,\ 1)$ (pour plus de précision à ce sujet voir [MAD02] et les références qui s'y trouvent). On

approxime alors $\mathbb{E}[\varphi(X)]$ par

$$S_N = \frac{1}{N}[\varphi(X_1) + \cdots + \varphi(X_N)].$$

C'est le principe des méthodes de Monte Carlo.

1.3 Résultats de Convergence

Nous présentons brièvement ici les différents résultats de convergence qui constituent en pratique la mise en œuvre des méthodes de Monte carlo. Ces résultats se trouvent dans tous les traités classiques de probabilité.

1.3.1 Loi Forte des Grands Nombres et méthode de Monte Carlo

Théorème .1. *Si on suppose que* $\mathbb{E}[\|\varphi(X)\|] < +\infty$ *et qu'on dispose d'une suite* $(X_n,\, n \geq 1)$ *de copies indépendantes de* X *alors, avec probabilité 1 on a :*

$$\lim_{n\to+\infty} \frac{1}{n}[\varphi(X_1) + \cdots + \varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

Dans la plupart des cas on dispose d'une estimation

$$\varepsilon_n = \mathbb{E}[\varphi(X)] - \frac{1}{n}[\varphi(X_1) + \cdots + \varphi(X_n)]$$

de l'erreur de convergence.

1.3.2 Théorème Central Limite et méthode de Monte Carlo

Théorème .2. *Si on suppose que* $\mathbb{E}[\|\varphi(X)\|^2] < +\infty$ *et si on note* σ^2 *la variance de* $\varphi(X)$*, alors :*

$$\frac{\sqrt{n}}{\sigma}\varepsilon_n \quad \textit{converge en loi vers} \quad G,$$

G étant une variable aléatoire suivant une loi normale centrée réduite.

Une conséquence (entre autres) de ce théorème est que pour tout $a < b$:

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\sigma}{\sqrt{n}}a \leq \varepsilon_n \leq \frac{\sigma}{\sqrt{n}}b\right) = \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

En pratique une façon standard de mesurer l'erreur de convergence d'une méthode de Monte Carlo est de donner soit l'écart-type $\frac{\sigma}{\sqrt{n}}$ de ε_n soit de donner un intervalle auquel appartient le résultat recherché avec une confiance (par exemple) de 95%. Cela conduit, étant donné que

$$\mathbb{P}(|G| \leq 1.96) \simeq 0.95,$$

à l'intervalle de confiance suivant :

$$\left[S_N - 1.96\frac{\sigma}{\sqrt{N}},\ S_N + 1.96\frac{\sigma}{\sqrt{N}}\right].$$

1.3.3 Estimation statistique de l'erreur ε_n

Les considérations précédentes prouvent que l'efficacité (ou la fiabilité des résultats) d'une méthode de Monte Carlo dépend de l'ordre de grandeur de la variance σ^2 de la variable aléatoire dont on veut estimer la moyenne. On reprend les notations précédentes. On rappelle que l'objectif est d'évaluer:

$$c = \mathbb{E}[\varphi(X)].$$

Soit $X_1, X_2, \ldots$ des réalisations indépendantes de la loi de X. D'après la Loi Forte des Grands Nombres on peut approcher c par $\hat{c}$ donné par:

$$S_N = \frac{1}{N}\sum_{n=1}^{N} \varphi(X_n),$$

pour N "suffisamment grand". Toujours d'après la Loi Forte des Grands Nombres et dans le cadre des hypothèses du Théorème Central Limite, il est bien connu que la quantité

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(\varphi(X_n) - S_N)^2$$

est un estimateur sans biais de la variance σ^2, appelée variance empirique de l'échantillon. On peut alors obtenir un intervalle de confiance approché (par exemple) à 95% en remplaçant dans l'intervalle de confiance fourni par le Théorème Central Limite σ par $\hat{\sigma}$

$$\left[S_N - 2\frac{\hat{\sigma}}{\sqrt{N}},\ S_N + 2\frac{\hat{\sigma}}{\sqrt{N}}\right].$$

Pour finir on peut dire que les avantages de cette méthode sont:

- sa progammation est facile,

- elle auto-estime son erreur d'approximation (sans pratiquement aucun calcul supplémentaire),
- elle ne dépend pas de la régularité de la fonction φ,
- et surtout, elle est "presque" insensible à la dimension d du problème.

Son inconvénient principal est:

- sa vitesse de convergence assez faible (de l'ordre de $\frac{C}{\sqrt{N}}$ avec $C > 0$).

Amélioration de la méthode de Monte Carlo standard

Pour améliorer la convergence d'une méthode de Monte carlo, l'on peut soit augmenter le nombre total de simulations N soit réduire la variance σ^2. Si l'on opte pour une augmentation de N, pour obtenir une précision supplémentaire d'un point l'on doit multiplier le nombre total de simulations par un facteur de 100 (ce qui devient très vite irréalisable compte tenu de la limitation de la puissance des machines). Si l'on opte pour une réduction de la valeur de σ, une réduction d'un facteur de 10 signifie par exemple que le nombre total de simulations requises pour obtenir le même degré de précision (que pour la méthode dans laquelle σ n'a pas été réduit) doit être diviser par 100. C'est ce qui justifie la très grande importance qui est accordée aux techniques, dites de réduction de variance, qui cherchent à diminuer la valeur de σ. Il existe une grande variété de ces techniques souvent à la fois simples et efficaces, mais dont le choix est spécifique au problème traité. Les plus connues (et aussi les plus classiques) sont

- La technique des variables de contrôle,
- La technique des variables antithétiques,
- La méthode de stratification,
- La méthode de conditionnement et
- La méthode des fonctions d'importance.

La dernière méthode appelée encore méthode d'échantillonnage préférentiel consiste à écrire

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left[\frac{\varphi(Y)f(Y)}{\bar{f}(Y)}\right],$$

où X et Y ont respectivement les densités $f(x)dx$ et $\bar{f}(y)dy$. Si l'on pose $\tilde{f}(Y) = \varphi(Y)f(Y)/\bar{f}(Y)$, on aura amélioré la méthode de Monte Carlo standard si

$$\mathbb{V}\mathrm{ar}(\tilde{f}(Y)) < \mathbb{V}\mathrm{ar}(\varphi(X)).$$

Les méthodes de réduction de variance que nous développons dans cette thèse, utilisent une version paramétrique de la technique d'échantillonnage préférentiel. Pour plus de détail sur les autres méthodes de réduction de variance nous citons entre autres les livres [LPS98] et [MAD02].

1.4 Martingales et Théorèmes limites

Après avoir défini la notion de martingale, nous exposons brièvement dans cette section quelques résultats probabilistes importants sur les martingales. Plusieurs livres fondamentaux exposent en détail, ces notions. Nous citons entre autres les livres de Williams [WIL95] (pour une initiation), de Neveu [NEV72] et de Hall-Heyde [HH80].

1.4.1 Définition et Propriétés

Définition .1. Soit $(M_n)_{n\geq 0}$ une suite de variables aléatoires adaptée à une filtration $(\mathcal{F}_n)_{n\geq 0}$.
On dit que $(M_n)_{n\geq 0}$ est une martingale par rapport à $(\mathcal{F}_n)_{n\geq 0}$ si pour tout $n \geq 0$, *p.s.*:

$$\mathbb{E}(M_{n+1}/\mathcal{F}_n) = M_n.$$

Remarque .1. Intuitivement, une martingale est une suite aléatoire qui a *tendance à rester constante en moyenne.*

Remarque .2. On rappelle qu'une espérance conditionnelle n'étant définie que presque sûrement, toutes les relations où apparaissent des espérances conditionnelles sont des relations presque sûres (*p.s.* en abrégé). Ainsi, même quand on en fera pas mention explicitement, elle sera sous-entendue.

1.4.2 Théorèmes Limites

Le théorème suivant qui nous sera utile dans la suite est relatif à la convergence des martingales de puissance $a-$intégrable avec $1 \leq a \leq 2$. On peut notamment trouver une preuve de ce théorème dans le livre de Hall et Heyde cité plus haut.

Théorème .3. *Soit $(M_n)_{n\geq 0}$ une martingale vectorielle de puissance $a-$intégrable avec $1 \leq a \leq 2$ par rapport à une filtration $(\mathcal{F}_n)_{n\geq 0}$. Notons*

$$\alpha_n = \mathbb{E}\left(\|M_{n+1} - M_n\|^a / \mathcal{F}_n\right) < +\infty, \quad et \quad S_n = \alpha_0 + \cdots + \alpha_n.$$

Alors on a

1. *Sur l'évènement* $\{S_\infty < \infty\}$, $M_n \longrightarrow M_\infty$ *où* M_∞ *est une variable aléatoire finie.*
2. *Sur l'évènement* $\{S_\infty = \infty\}$ *on a une loi des grands nombres,*

$$\frac{M_n}{S_n} \xrightarrow{p.s.} 0.$$

Remarque .3. Lorsque $a = 2$, $(S_{n-1})_{n\geq 1}$ est le *processus croissant* ou *crochet oblique* associé à la martingale de carré intégrable $(M_n)_{n\geq 0}$ et l'on note $\langle M\rangle_n = S_{n-1}$. Dans ce cas

$$\alpha_n = \mathbb{E}\left(\|M_{n+1} - M_n\|^2 / \mathcal{F}_n\right) = \mathbb{E}\left(\|M_{n+1}\|^2 / \mathcal{F}_n\right) - \|M_n\|^2 .$$

Remarque .4. La partie 1) du théorème est le Théorème de Chow classique. La partie 2) en résulte en considérant la nouvelle martingale (martingale transformée de $(M_n)_{n\geq 0}$)

$$N_n = \sum_{k=1}^{n} (M_k - M_{k-1})/(S_{k-1} b(\ln S_{k-1}))^{1/a},$$

où b est une fonction vérifiant $\int_1^{+\infty} (1/b(t))dt < +\infty$, et en lui appliquant le résultat de la partie 1) et le Lemme de Kronecker (que nous rappelons à la fin de ce chapitre).

Théorème .4. *Soit* $(M_n)_{n\geq 0}$ *une martingale (vectorielle) par rapport à une filtration* $(\mathcal{F}_n)_{n\geq 0}$*, de carré intégrable pour laquelle il existe une suite positive déterministe* $(v_n)_{n\geq 0}$ *croissant vers* $+\infty$ *tel que son processus crochet* $\langle\ M\ \rangle_n$ *vérifie:*

$$(1) \qquad \frac{\langle\ M\ \rangle_n}{v_n} \xrightarrow[n]{\mathbb{P}} \Sigma^2 \quad \textit{avec} \quad \Sigma^2 \quad \textit{une matrice carrée définie positive;}$$

alors on a la loi forte des grands nombres suivante

$$\frac{M_n}{v_n} \xrightarrow[n]{p.s.} 0.$$

Si de plus la condition de Lindberg est vérifiée:

$$(2) \qquad \forall \varepsilon > 0, \quad \frac{1}{v_n} \sum_{k=1}^{n} \mathbb{E}\left[\|M_k - M_{k-1}\|^2 1_{\{\|M_k - M_{k-1}\| \geq \varepsilon \sqrt{v_k}\}} / \mathcal{F}_{k-1} \right] \xrightarrow[n]{\mathbb{P}} 0,$$

alors:

$$\frac{M_n}{\sqrt{v_n}} \xrightarrow[n]{\mathcal{L}} \mathcal{N}(0, \Sigma^2).$$

Pour une preuve de ce résultat, l'on peut consulter [NEV72] ou [DD83].

1.5 Algorithmes Stochastiques et Projection à la Chen

Il n'est pas nécessaire de rappeler ici tous les domaines des sciences appliquées dans lesquels peuvent intervenir les algorithmes stochastiques. Cela va du calcul probabiliste des structures (Mécanique) à l'étude des réseaux de neurones en passant par l'automatique, l'économie ou encore la chimie. L'utilisation de ces algorithmes dans les problèmes de la finance numérique est peu développée. La recherche des points où une fonction réelle V atteint ses minima (avec sa transposition en recherche des maxima pour la fonction $-V$)

$$\operatorname{argmin} V = \left\{ x,\, V(x) = \inf_y \{V(y)\} \right\},$$

est un problème très courant qu'on ramène souvent à celui de la recherche de l'ensemble des points où une fonction h s'annule

$$h^{-1}(0) = \{x, h(x) = 0\},$$

avec $h = \nabla V$, le gradient de V. Le caractère stochastique des procedures de recherche de zéro de h provient souvent de ce que

- ou bien les fonctions h et V ne sont observables qu'à des perturbations près,
- ou bien h est difficile à expliciter, mais on sait facilement simuler une quantité aléatoire qui lui est proche "en moyenne".

Il arrive aussi que ce caractère stochastique provienne d'un bruit aléatoire que l'on utilise pour interdire à l'algorithme de converger vers des solutions aberrantes (minima locaux, points selles, ...). La forme générale des algorithmes stochastiques pour le problème de recherche de zéro de h s'écrit

$$\theta_{n+1} = \theta_n - \gamma_{n+1}[h(\theta_n) + \varepsilon_{n+1}] + \sigma_{n+1}\eta_{n+1}. \tag{.1.1}$$

La fonction h est définie sur un ouvert $\mathcal{O}$ de $\mathbb{R}^d$, à valeurs dans $\mathbb{R}^d$. Les pas (ou gains) de l'algorithme $(\gamma_n)_{n\geq 0}$ et $(\sigma_n)_{n\geq 0}$ sont deux suites déterministes, de nombres réels strictement positifs et décroissants (au sens large) qu'on peut régler à priori librement. Les suites $(\varepsilon_n)_{n\geq 0}$ et $(\eta_n)_{n\geq 0}$ sont des perturbations aléatoires imposées par le modèle. Par exemple tandis que $(\varepsilon_n)_{n\geq 0}$ sera le plus souvent un accroissement de martingale (*i.e.* un bruit blanc), $(\eta_n)_{n\geq 0}$ sera une "petite perturbation" éventuellement rajoutée pour les besoins de convergence.

Nous renvoyons aux livres de Duflo Marie [DUF96] et [DUF97], à celui de Benveniste, Metivier et Priouret [BMP87], à celui de Pflug [PFL96] ou encore au livre très récent de Kushner et Yin [KY03] pour un exposé détaillé sur le sujet. Il existe également une multitude d'articles sur les propriétés de convergence de tels algorithmes dont on pourra trouver certains dans les références bibliographiques à la fin de ce manuscrit.
Dans les cas qui vont nous occuper tout au long de ce travail, l'algorithme considéré pourra toujours se mettre sous la forme

$$\theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1}\varepsilon_{n+1} + \gamma_{n+1} r_{n+1},$$

avec $\mathbb{E}[\varepsilon_{n+1}/\mathcal{F}_n] = 0$, $(\mathcal{F}_n)_{n\geq 0}$ étant une filtration engendrée par la suite $(\theta_n)_{n\geq 0}$. $(r_n)_{n\geq 0}$ est une suite de termes résiduels qu'il faudra controler.
Dans la suite de cette section, nous présentons plus en détail les algorithmes stochastiques (markoviens sur $\mathbb{R}^d$) et quelques outils permettant *l'analyse numérique de leur comportement en temps long* les résultats rappelés ici nous servirons dans les prochains chapitres de cette thèse.

1.5.1 Présentation générale

On considère une suite de vecteurs aléatoires de $\mathbb{R}^d$ définie de façon récursive par

$$\theta_0 \quad \text{donné}, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} F(\theta_n, X_{n+1}), \quad n \geq 0. \tag{.1.2}$$

- $(\gamma_n)_{n\geq 0}$ est une suite (décroissante au sens large) de réels positifs , appelée *pas* ou *gain* de l'algorithme stochastique, vérifiant

$$\gamma_n > 0 \quad \text{et} \quad \sum_{n\geq 0} \gamma_n = +\infty.$$

- $(X_n)_{n\geq 0}$ est une suite de vecteurs aléatoires indépendants suivant tous la même loi que X.

- La fonction $F(\cdot)$ est *l'observation* de l'algorithme stochastique. C'est un champ de vecteurs défini sur $\mathbb{R}^d \times \mathbb{R}^q$ à valeurs dans $\mathbb{R}^d$ et qui vérifie

$$\forall \theta \in \mathbb{R}^d, \mathbb{E}[F(\theta, X)] < +\infty, \quad \text{et} \quad \theta \mapsto F(\theta, \cdot) \quad \text{continue} \quad d\mathbb{P}_X - p.s.$$

La fonction h donnée par

$$h(\theta) = \mathbb{E}[F(\theta, X)], \tag{$\star$}$$

est appelée *fonction moyenne* de l'algorithme. La fonction h est en général numériquement difficile d'accès. Notons que l'on a

$$\mathbb{E}[F(\theta_n, X_{n+1})/\mathcal{F}_n] = h(\theta_n), \tag{.1.3}$$

avec $\mathcal{F}_n = \sigma\{\theta_0, X_i, 0 \leq i \leq n\}$. $(\mathcal{F}_n)_{n\geq 0}$ est la filtration propre de $(X_n)_{n\geq 0}$ (et de θ_0). On peut réecrire (.1.2) sous la forme

$$\theta_{n+1} = \theta_n - \gamma_{n+1}h(\theta_n) + \gamma_{n+1}(h(\theta_n) - F(\theta_n, X_{n+1})), \quad n \geq 0. \tag{.1.4}$$

Les termes $h(\theta_n) - F(\theta_n, X_{n+1}), \quad n \geq 0$ ont une espérance conditionnelle nulle par rapport à $\mathcal{F}_n$ et sont par ailleurs mésurables par rapport à $\mathcal{F}_{n+1}$. Ce sont donc les accroissements d'une $\mathcal{F}_n-$martingale $(M_n)_{n\geq 0}$. Finalement l'algorithme stochastique (.1.2) s'écrit

$$\theta_{n+1} = \theta_n - \gamma_{n+1}h(\theta_n) + \gamma_{n+1}\Delta M_{n+1}, \quad n \geq 0. \tag{.1.5}$$

Les outils utilisés pour étudier cet algorithme diffèrent selon que le pas $(\gamma_n)_{n\geq 0}$ est constant ou tend vers 0.

- $\gamma_n = \gamma > 0$:
  - L'algorithme (.1.2) est alors une $\mathcal{F}_n-$chaîne de Markov homogène de semi-groupe de transition *fellérien* donnée par

    $$Q_\gamma(f)(\theta) := \mathbb{E}[f(\theta - \gamma F(\theta, X))].$$

  - On peut également voir cet algorithme, via (.1.5), comme un schéma d'Euler de pas γ de l'équation différentielle ordinaire associée à la fonction $(-h)$, à savoir

    $$\dot{\theta}(t) = -h(\theta(t)) \quad t \geq 0, \tag{ODE_h}$$

    auquel l'on a rajouté un bruit $\gamma\Delta M_{n+1}$.

  Dans ce cas, pour γ suffisamment petit, le comportement de l'algorithme stochastique (.1.2) est très proche de celui de la solution de (ODE_h).

- $\gamma_n \to 0$:
  - Il s'agit alors d'une chaîne de Markov inhomogène, dont la transition à la date n est donnée par

    $$\mathbb{E}[f(\theta_{n+1})/\mathcal{F}_n] = Q_{\gamma_{n+1}}f(\theta_n).$$

– Ou, via (.1.5), le voir comme un schéma d'Euler *à pas "décroissant"* et *perturbé* de (ODE_h).

Dans ce dernier cas l'algorithme stochastique (.1.2) permet d'estimer asymptôtiquement la solution de l'équation

$$h(\theta) = 0, \quad \theta \in \mathbb{R}^d, \tag{$\star\star$}$$

où h est définie par $(\star)$.

Il existe une littérature abondante sur *l'analyse numérique du comportement en temps long* de l'algorithme stochastique (.1.2). Nous allons simplement rappelé ici quelques résultats de convergence qui servirons dans les prochains chapitres. Dans la suite on utilisera la notation $Y_{n+1} = F(\theta_n, X_{n+1}, \quad n \geq 0$.

1.5.2 Le cas où le pas $(\gamma_n)_{n\geq 0}$ décroit vers 0

L'une des difficultés dans l'utilisation pratique des Approximations Stochastiques réside dans le fait que leurs versions diffèrent sensiblement d'un domaine d'applications à un autre. De fait, les critères théoriques de convergence sont tout aussi nombreux mais pas toujours évidents à adapter à la situation réelle de l'utilisateur. On peut trouver une démonstration du théorème ci-dessous dans [DUF97].

Théorème .5. *Sous les hypothèses suivantes*

$$(H_1) \qquad \exists\, \theta^* \in \mathbb{R}^d,\ h(\theta^*) = 0, \quad \forall \theta \in \mathbb{R}^d\ \theta \neq \theta^*, \quad (\theta - \theta^*) \cdot h(\theta) > 0, \tag{.1.6}$$

$$(H_2) \qquad \sum_n \gamma_n = +\infty \quad et \quad \sum_n \gamma_n^2 < +\infty, \tag{.1.7}$$

$$(H_3) \qquad \exists\, C > 0, \quad \mathbb{E}[\|Y_{n+1}\|^2/\mathcal{F}_n] < C(1 + \|\theta_n\|^2) \quad p.s., \tag{.1.8}$$

l'algorithme (.1.2) $(\theta_n)_{n\geq 0}$ *converge presque sûrement vers* θ^*.

Remarque .5. Un point intéressant à noter est que la convergence de la suite $(\theta_n)_{n\geq 0}$ ne dépend pas de la valeur initiale θ_0.

Remarque .6. Pour une présentation très succinte des algorithmes stochastiques, l'on peut consulter [MD02]. Pour une étude plus appronfondie l'on se reportera par exemple à des livres généraux comme [PFL96], [BMP87], [DUF97] ou [DUF96]. Pour des discussions plus spécialisées il existe une multitude d'articles comme par exemple [CLA84], [CZ86], [DEL96] ou [CGG88].

Remarque .7. De toutes les hypothèses du théorème .5, (H_3) est de loin celle qui pose le plus de difficulté. L'hypothèse (H_1) est ce qu'on appelle une condition de *type Lyapounov*. Elle traduit une force de rappel et assure la stabilité de (ODE_h). Nous reviendrons plus en détail sur cette notion plus tard.

Dans le théorème ci-dessus, si on remplace la condition (H_3) par

$$(H_4) \qquad \sum_{n=0}^{+\infty} \gamma_{n+1}^2 \mathbb{E}[\|Y_{n+1}\|^2/\mathcal{F}_n] \leq c < +\infty \qquad p.s., \tag{.1.9}$$

le résultat reste inchangé (voir [BLL86]).

1.5.3 Méthode de Projection à la "Chen"

En pratique il est difficile de montrer que ces conditions sont vérifiées comme l'illustre l'exemple unidimensionnel très simple suivant (emprunté à [CHE02])

$$F(\theta, x) = -(\theta - 10)^3 \quad \text{avec} \quad \gamma_n = \frac{1}{n}.$$

Il est clair que dans ce cas l'hypothèse (H_3) du théorème .5 n'est pas satisfaite. De plus il n'est pas évident de montrer que (H_4) est vérifiée.
D'ailleurs quelques itérations donnent à partir de la valeur initiale $\theta_0 = 0$

$$\theta_1 = 1000, \quad \theta_2 = -485148500, \quad \theta_3 \simeq 3.8 \times 10^{25},$$

suggérant du même coup que l'algorithme (dans ce cas) diverge très rapidement, et la nécessité d'avoir une hypothèse d'un type (H_3) ou (H_4) mentioné ci-dessus. Toutefois lorsqu'on choisit la valeur initiale θ_0 de l'algorithme de façon que $|\theta_0 - 10| < 1$, alors ce dernier converge vers la solution du problème qui est manifestement $\theta^* = 10$. Chen observe à propos de cet exemple que choisir la valeur initiale θ_0 dans un voisinage proche de 10, est en un sens équivalent à considérer les itérations de l'algorithme $(\theta_n)_{n\geq 0}$ à partir d'un certain rang n_0. On comprend bien que la difficulté réside dans le choix de ce rang n_0. Pour contourner le problème, Chen H-F. (1986) propose une méthode de *projections aléatoires* de l'agorithme. La méthode est la suivante:
D'abord on généralise l'hypothèse de Lyapounov (H_1)

(H_5) $\exists\, \theta^* \in \mathbb{R}^d, \quad \exists\, V : \mathbb{R}^d \to \mathbb{R},$ deux fois continûment différentiable tels que,

$$V(\theta^*) = 0,\ \lim_{\|x\|\to\infty} V(x) = +\infty, \text{et } \forall\ \theta \neq \theta^*\ V(\theta) > 0,\ h(\theta) \cdot \nabla V(\theta) > 0.$$

Remarque .8. A partir de cette hypothèse on retrouve en effet (H_1) avec la fonction de Lyapounov $V(\theta) = \frac{1}{2}\|\theta - \theta^*\|^2$.

Ensuite, on choisit deux vecteurs distincts θ^1 et θ^2 de $\mathbb{R}^d$ et on fixe un réel $M > 0$ tels que

$$\max(V(\theta^1), V(\theta^2)) < \min(M, \inf(V(\theta); \|\theta\| > M)). \tag{.1.10}$$

Remarque .9. Une fois $\theta^1 \neq \theta^2$ fixés, le choix de M est toujours possible puisque $\lim_{\|x\|\to\infty} V(x) = +\infty$.

Finalement on considère une suite quelconque $(U_n)_n$ de réels positifs strictement croissante, tendant vers l'infini avec $U_0 >> M$ (dans leur article, les auteurs choisisent $U_0 \geq M + 8$ pour les besoins techniques de la preuve qu'ils donnent de leur principal résultat), pour $n \geq 0$ et θ_0 donné, on définit par récurrence

$$\theta_{n+1} = \begin{cases} \theta_n - \gamma_{n+1} Y_{n+1} & \text{si} \quad \|\theta_n - \gamma_{n+1} Y_{n+1}\| \leq U_{\sigma(n)}, \\ \bar{\theta}_n & \text{sinon} \end{cases} \tag{.1.11}$$

avec

$$\sigma(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{\|\theta_k - \gamma_{k+1} Y_{k+1}\| > U_{\sigma(k)}\}}, \quad \sigma(0) = 0, \tag{.1.12}$$

et

$$\bar{\theta}_n = \begin{cases} \theta^1 & \text{si } \sigma(n) \text{ est pair,} \\ \theta^2 & \text{si } \sigma(n) \text{ est impair.} \end{cases} \tag{.1.13}$$

Voici comment on peut interpréter cette technique de projection. Alors que $\sigma(n)$ désigne le nombre de projections effectuées aucours des n premières itérations, $U_{\sigma(n)}$ représente la taille maximale que ne doit pas dépasser l'algorithme au bout de $(n+1)$ itérations. En effet on voit de (.1.11) que si l'estimé de rang $(n+1)$ calculer par l'algorithme de Robbins–Monro reste dans la sphère fermée de rayon aléatoire $U_{\sigma(n)}$ alors il évolue exactement comme un algorithme de RM classique (.1.2). Par contre si $(\theta_n - \gamma_{n+1} Y_{n+1})$ sort de cette région alors l'estimé à la date $(n+1)$ est réinitialisé à un point fixé dans le passé, et la limite de projection est dilatée de $U_{\sigma(n)}$ à $U_{\sigma(n)+1}$. Dans cette méthode, on n'impose pas à l'algorithme θ_n d'être à priori borné. C'est d'ailleurs le trait le plus remarquable de cette méthode de projection dûe Chen.

Remarque .10. Il faut noter au passage que l'algorithme (.1.11) peut prendre la forme (.1.1), avec $\gamma_n = \sigma_n$, $\varepsilon_{n+1} = h(\theta_n) - Y_{n+1}$ et $\eta_{n+1} = (\bar{\theta}_n - \theta_n + \gamma_{n+1} Y_{n+1}) \times \mathbf{1}_{\{}\|\theta_n - \gamma_{n+1} Y_{n+1}\| > U_{\sigma(n)}\}$

Le théorème suivant est prouvé dans [CGG88].

Théorème .6. *Supposons que l'hypothèse* (H_5) *est vérifiée. Si la condition suivante est satisfaite*

$$(H_6) \qquad \overline{\lim_{n\to+\infty}} \gamma_{n+1} \| \sum_{i=0}^{n-1} \Delta M_{i+1} \| = 0 \qquad p.s., \qquad (.1.14)$$

alors l'algorithme de Robbins-Monro tronqué par la méthode de Chen (.1.11) *converge presque sûrement vers* θ^* *et le nombre de projections* σ_n *est presque sûrement borné.*

Ce résultat sera utilisé dans le prochain chapitre.

1.5.4 Méthode de Projection fixe

Comme on l'a vu dans l'exemple de la section précédente les itérations d'algorithmes stochastiques peuvent exploser très vite. Cela peut provenir d'un mauvais choix de la structure de l'algorithme dans la modélisation d'un phénomène physique (réel) ou tout simplement d'un artefact numérique. Un aspect important des applications des algorithmes stochastiques concerne donc l'attitude à adopter lorsque les valeurs des itérations deviennent trop grandes. Le problème est qu'il n'existe pas de procédure universelle permettant de résoudre ce genre de situation. Néanmoins dans la plupart des situations pratiques, les problèmes rencontrés possèdent, en réalité, des bornes implicites (contraintes physiques ou économiques). Il convient alors, en cas de valeurs excessives, de tronquer les itérations de l'algorithme. Il faut mentionner d'ailleurs que même dans le cas où l'on a démontré qu'un algorithme stochastique (non borné *à priori*) converge fortement, il est très courant que pour certains jeux de paramètres, ses itérations successives explosent numériquement. En général, dans l'étude mathématique de ces algorithmes, on remplace les hypothèses de "bornitudes des itérations successives" par diverses hypothèses de stabilité. Alors évidemment cela aboutit à des développements mathématiques compliqués (beaucoup plus compliqués que dans le cas borné) et intéressants (sur le plan de la théorie, voir par exemple, [PAG03]), mais en pratique, cela n'empêche pas de devoir tronquer les itérations successives de l'algorithme lorsqu'il s'agit de le mettre en œuvre numériquement.
Soit K un compact non vide de $\mathbb{R}^d$. Un algorithme stochastique $(\theta_n)_{n\geq 0}$ projeté sur K sera défini récursivement par la donnée de $\theta_0 \in K$ et par

$$\theta_{n+1} = \Pi_K(\theta_n - \gamma_{n+1} Y_{n+1}). \qquad (.1.15)$$

On peut réecrire cet algorithme sous la forme

$$\theta_{n+1} = \theta_n - \gamma_{n+1} Y_{n+1} + \gamma_{n+1} Z_{n+1}, \qquad (.1.16)$$

où $(Z_n)_{n\geq 1}$ est un terme de correction ou de projection sur K. En faisant apparaître les observations moyennes $h(\theta_n) = \mathbb{E}[Y_{n+1}/\mathcal{F}_n]$, avec $\mathcal{F}_n = \sigma\{Y_k,\, k \leq n\}$ cet algorithme lui même peut être réecrit sous la forme

$$\theta_{n+1} = \theta_n + \gamma_{n+1} h(\theta_n) + \gamma_{n+1}\varepsilon_{n+1} + \gamma_{n+1} Z_{n+1}, \tag{.1.17}$$

où $\varepsilon_{n+1} = -Y_{n+1} + h(\theta_n)$. Le résultat de convergence suivant sera utilisé dans le Chapitre 3 et est dû à Bernard Delyon 2000 (voir [DEL00]).

Théorème .7. *Si*

- $\lim_{n\to+\infty} \sum_{i=1}^{n} \gamma_i \varepsilon_i < +\infty \quad et \quad \lim_{n\to+\infty} \|Z_n\| = 0, \quad p.s.$
- *Il existe* $\theta^* \in \overset{\circ}{K}$ *et une fonction* $V : \mathbb{R}^d \to \mathbb{R}_+$, *convexe, de classe* $\mathcal{C}^1$ *tels que*

$$\lim_{\|x\|\to+\infty} V(x) = +\infty, \quad V(\theta^*) = 0 \quad et \quad \nabla V(\theta)\cdot h(\theta) > 0, \forall\, \theta \neq \theta^*,$$

alors l'algorithme stochastique projeté $(\theta_n)_{n\geq 0}$ *converge presque sûrement vers* θ^*.

1.5.5 Le cas des Algorithmes Stochastiques à gain constant

Toute cette partie constitue un condensé de résultats connus. Pour la cohérence de notre exposé, il nous a semblé important de faire un rappel sommaire de ces résultats. Ils sont essentiellement tirés du livre de Kushner et Yin (2003 [KY03]).
Nous avons vu à la section 1.5.1 que les algorithmes à pas constants s'écrivaient

$$\theta^\gamma_{n+1} = \theta^\gamma_n - \gamma h(\theta^\gamma_n) + \gamma \varepsilon^\gamma_{n+1}, \quad n \geq 0, \quad \theta^\gamma_0 = \theta_0, \tag{.1.18}$$

avec $h(\theta^\gamma_n) = \mathbb{E}\left[Y^\gamma_{n+1}/\mathcal{F}^\gamma_n\right]$, Y^γ_{n+1} l'observation de l'algorithme et $\varepsilon^\gamma_{n+1} = h(\theta^\gamma_n) - Y^\gamma_{n+1}$. Lorsque le pas $\gamma > 0$ est suffisamment petit, on peut montrer que pour n assez grand la loi de θ^γ_n ressemble à une gaussienne de moyenne θ^* et de matrice de variance-covariance $\sqrt{\gamma}\,\Sigma$, Σ étant une matrice définie positive indépendante de γ. En d'autres termes pour γ assez petit, si θ^γ désigne la limite en loi de (.1.18) alors

$$\theta^\gamma \overset{\mathcal{L}}{\simeq} \theta^* + \sqrt{\gamma}\, G, \quad \text{avec} \quad G \sim \mathcal{N}(0, \Sigma).$$

Ainsi moralement lorsque γ devient petit, la loi de (.1.18) va se concentrer autour de θ^*. On rappelle que θ^* est solution de l'équation

$$h(\theta) = 0, \quad \theta \in \mathbb{R}^d. \tag{.1.19}$$

On va montrer que sous des conditions très simples, θ^* est aussi limite d'une solution de l'équation différentielle ordinaire (EDO en abrégé)

$$\dot{\theta}(t) = -h(\theta(t)), \quad \theta(0) = \theta_0. \tag{EDO_h}$$

La technique la plus couramment utilisée pour étudier la stabilité des solutions des EDO est la méthode de Lyapounov.

1.5.6 Méthode de Lyapounov pour les EDO

Nous commençons par la définition des fonctions de Lyapounov

Définition .2. Une fonction $V : \mathbb{R}^d \to \mathbb{R}_+$ est dite de Lyapounov si elle est convexe, de classe $\mathcal{C}^1$, et vérifie $\lim_{\|x\| \to +\infty} V(x) = +\infty$.

Le principe de la méthode de Lyapounov est le suivant (on peut se repporter à [LL61]). Soit h une fonction de $\mathbb{R}^d$ dans lui même et $\theta(\cdot)$ une solution de l'EDO (.1.19). On suppose qu'il existe une fonction de Lyapounov V telle que:

$(A4.1) \quad \exists \hat{\theta}, \in \mathbb{R}^d \quad V(\hat{\theta}) = 0 \quad \text{et} \quad V(\theta) > 0 \, \forall \, \theta \neq \hat{\theta}.$

et

$(A4.2) \quad \nabla V(x) \cdot h(x) > 0 \quad \forall x \in \mathbb{R}^d - \{\hat{\theta}\}.$

Pour $\lambda > 0$ si on pose $\Gamma_\lambda = \{x; V(x) \leq \lambda\}$ alors il est clair que

$$\begin{aligned} \frac{d}{dt}\left[V(\theta(t))\right] =& \nabla V(\theta(t)) \cdot \dot{\theta}(t) \\ =& - \nabla V(\theta(t)) \cdot h(\theta(t)) \\ :=& - k(\theta(t)) \\ \leq& 0 \quad \forall \, t \geq 0, \end{aligned}$$

et la fonction $t \mapsto V(\theta(t))$ est décroissante, d'où la proposition suivante

Proposition 1.5.1. *Soit $\theta(\cdot)$ une solution de l'EDO (.1.19) de condition initiale θ_0. Sous les hypothèses (A4.1) et (A4.2), on a*

$$\theta_0 \in \Gamma_\lambda, \Rightarrow \theta(t) \in \Gamma_\lambda, \quad \forall \, t > 0.$$

Maintenant comme la fonction V est positive on a

$$\begin{aligned} V(\theta(0)) &\geq V(\theta(0)) - V(\theta(t)) \\ &= -\int_0^t \frac{d}{dt}\left[V(\theta(s))\right] ds \\ &= \int_0^t k(\theta(s)) ds. \end{aligned}$$

L'application $x \mapsto k(x)$ étant elle même positive (et continue), on a

$$0 \leq \int_0^{+\infty} k(\theta(t)) dt \leq V(\theta(0)) < +\infty, \quad \text{et} \quad \lim_{t\to+\infty} k(\theta(t)) = 0.$$

Ce qui donne la proposition suivante

Proposition 1.5.2. *Soit $\theta(\cdot)$ une solution de l'EDO* (.1.19) *de condition initiale θ_0. Sous les hypothèses* $(A4.1)$ *et* $(A4.2)$*, si $\lim_t \theta(t)$ désigne une valeur d'adhérence de $(\theta(t))_{t\geq 0}$, alors*

$$\lim_t \theta(t) \in \{x \in \Gamma_\lambda;\, k(x) = 0\},$$

$\lambda > 0$ étant choisi tel que $\theta(0) \in \Gamma_\lambda$.

Remarque .11. Si la fonction k est seulement mesurable mais est telle que

$$\forall\, \delta > 0, \exists\, \varepsilon > 0 \quad t.q. \quad |x| \geq \delta \Rightarrow k(x) \geq \varepsilon,$$

alors la conclusion de la proposition 1.5.2 est encore vraie.

Remarque .12. La méthode de Lyapounov pour la stabilité des EDO fonctionne bien lorsque la fonction k s'annule en des points isolés. Lorsque ce n'est le cas, il convient d'utiliser un résultat appelé *Théorème d'invariance de LaSalle* (voir [LL61]) pour s'en sortir. Dans tous les cas que nous allons rencontrer, la fonction k s'annule en un seul et unique point noté θ^* et ce problème ne se pose donc pas.

Maintenant nous allons de revisiter un résultat de convergence des algorithmes stochastiques à pas constant. En fait nous nous contenterons de définir le cadre et de présenter un premier résultat de convergence.

1.5.7 Un résultat général de convergence faible

Le résultat de convergence que nous énonçons ici est tiré de [KY03]. On considère l'algorithme stochastique *projeté* suivant:

$$\theta_{n+1}^\gamma = \Pi_K(\theta_n^\gamma - \gamma Y_{n+1}^\gamma),\, n \geq 0, \quad \theta_0^\gamma \in K, \tag{.1.20}$$

où K est un compact non vide de $\mathbb{R}^d$. En pratique, le plus souvent le compact K est :

- Soit un hypercube fermé de $\mathbb{R}^d$,
- Soit $K = \{x;\, q_i(x) \leq 0;\, i = 1, \dots p\}$ où $p \geq 1$ pour $i = 1, \dots p$, $q_i(\cdot)$ est une fonction réelle de classe $\mathcal{C}^1$ définie sur $\mathbb{R}^d$ telle que $\nabla q_i(x) \neq 0$ si $q_i(x) = 0$.

Le second cas généralise le premier et est l'exemple le plus couramment rencontré dans les applications.

Remarque .13. Dans le second cas, il est clair que K est une sous-variété, connexe, compact et non vide de $\mathbb{R}^d$.

Les q_i sont alors ce qu'on appelle les contraintes de l'algorithme. Une contrainte q_i est dite active en un point x si $q_i(x) = 0$. On pose $A(x) = \{i;\, q_i(x) = 0\}$ l'ensemble des indices des contraintes actives en x, et on note $C(x)$ le cône convexe engendré par $\{y;\, y = \nabla q_i(x);\, i \in A(x)\}$ l'ensemble des normales extérieurs en x.
Si on suppose que pour chaque x tel que $A(x) \neq \emptyset$, l'ensemble $\{\nabla q_i(x);\, i \in A(x)\}$ est un système libre (au sens vectoriel du terme) alors on peut montrer que $C(x) = \{0\}$ en tout point $x \in K$ tel que $q_i(x) \neq 0$.

Remarque .14. Il est facile de voir que si K est une boule fermée de $\mathbb{R}^d$ alors $C(x) = \{0\}$ en tout point $x \in \overset{\circ}{K}$ appartenant à l'interieur de K.

On peut toujours réecrire l'algorithme (.1.20) sous la forme

$$\theta_{n+1}^{\gamma} = \theta_n^{\gamma} - \gamma Y_{n+1}^{\gamma} + \gamma Z_{n+1}^{\gamma}, \quad \theta_0^{\gamma} \in K, \tag{.1.21}$$

où $(Z_n^{\gamma})_{n\geq 1}$ est un terme de correction ou de projection sur K. On pose $\mathcal{F}_n^{\gamma} = \sigma\{\theta_k^{\gamma}, Y_k^{\gamma};\, k \leq n\}$ la tribu engendrée par θ_k^{γ} et Y_k^{γ} pour $k \leq n$ et on fait les hypothèses suivantes:

(H_7) $\{Y_n^{\gamma};\, n, \gamma\}$ est uniformément intégrable.

(H_8) il existe des fonctions $h_n^{\gamma}(\cdot)$ uniformément continues en n et γ et des vecteurs aléatoires ε_n^{γ} tels que:$\mathbb{E}\left[Y_{n+1}^{\gamma}/\mathcal{F}_n^{\gamma}\right] = h_n^{\gamma}(\theta_n^{\gamma}) + \varepsilon_{n+1}^{\gamma}$.

(H_9) il existe une fonction continue $\bar{h}(\cdot)$ telle que $\forall \theta \in K$
$\lim_{m,n,\gamma} \frac{1}{m} \sum_{k=n}^{m+n-1} \left[h_n^{\gamma}(\theta) - \bar{h}(\theta)\right] = 0$

(H_{10}) $\lim_{m,n,\gamma} \frac{1}{m} \sum_{k=n}^{m+n-1} \mathbb{E}\left[\varepsilon_{k+1}^{\gamma}/\mathcal{F}_n^{\gamma}\right] = 0$ en moyenne.

Le symbole $\lim_{m,n,\gamma}$ signifie que la limite est considérée pour $m, n \to +\infty$ et $\gamma \to 0$ de quelque façon que ce soit. On définit les processus interpolés de l'algorithme (.1.21) de la manière suivante:
D'abord pour $t = 0$, $\theta^\gamma(0) = \theta_0^\gamma$ et $\theta^\gamma(t) = \theta_n^\gamma$ sur l'intervalle de temps $[n\gamma,\, n\gamma + \gamma[$. Ensuite $Z^\gamma(0) = 0$ et $Z^\gamma(t) = \gamma \sum_{k=0}^{[t/\gamma]} Z_k^\gamma$ pour $t > 0$. $[a]$ désigne la partie entière d'un réel a et par convention une somme sur un ensemble d'indice vide est nulle. Et enfin on définit de manière analogue le processus $Y^\gamma(\cdot)$.
Le théorème suivant est prouvé dans [KY03] (page 248).

Théorème .8. *Sous les hypothèses* (H_7)-(H_{10}), *pour toute famille croissante d'entiers* $(N_\gamma)_{\gamma>0}$, *pour toute sous-suite de* $\{\theta^\gamma(\gamma N_\gamma + \cdot),\, Z^\gamma(\gamma N_\gamma + \cdot),\, \gamma > 0\}$, *il existe une sous-suite que nous noterons encore* $\{\theta^\gamma(\gamma N_\gamma + \cdot),\, Z^\gamma(\gamma N_\gamma + \cdot),\, \gamma > 0\}$ *et un processus* $(\theta(\cdot),\, Z(\cdot))$ *tels que*

$$(\theta^\gamma(\gamma N_\gamma + \cdot),\, Z^\gamma(\gamma N_\gamma + \cdot)) \xrightarrow[\gamma\to 0]{\mathcal{L}} (\theta(\cdot),\, Z(\cdot))\,,$$

avec

$$\theta(t) = \theta(0) + \int_0^t h(\theta(s))ds + Z(t).$$

De plus il existe un processus intégrable $z(t)$ *tel que*

$$Z(t) = \int_0^t z(s)ds \quad et \quad z(t) \in -C(\theta(t)),$$

pour presque tout t *et* ω.
Lorsque $K = \mathbb{R}^d$ *(i.e. il n'y a pas de projection), si on fait l'hypothèse supplementaire que la suite* $(\theta_n^\gamma)_{\gamma>0}$ *est tendue alors les conclusions ci-dessus sont encore vraies avec* $Z(t) \equiv 0 \quad \forall\, t > 0$.

Remarque .15. Lorsque $N_\gamma = 0$, $\forall \gamma > 0$ et $\theta^\gamma(0) = \theta_0$, le résultat de la convergence faible du Théorème .8 ci-dessus nous dit que la trajectoire $\theta^\gamma(\cdot)$ est très proche de la trajectoire de la solution de l'EDO $\dot\theta(t) = \bar h(\theta(t))$; $\theta(0) = \theta_0$ avec une probabilité arbitrairement grande (uniformément par rapport a la condition initiale) quand $\gamma \to 0$.

Remarque .16. Moyennant quelques hypothèses supplémentaires, on peut préciser la vitesse de convergence faible du Théorème .8 ci-dessus (voir [KY03] page 319 ou [BMP87] chap.4). En effet on montre que si $\theta_0^\gamma = \theta_0$ alors

$$\frac{(\theta_n^\gamma - \theta_0(n\gamma))}{\sqrt{\gamma}} \xrightarrow[\gamma\to 0]{\mathcal{L}} \mathcal{N}(0,\, \Sigma),$$

Σ étant une matrice de variance-covariance (ne dépendant pas de γ) et $(\theta_0(t))_{t\geq 0}$ la solution de l'EDO $\dot\theta(t) = \bar h(\theta(t))$ avec comme condition initiale $\theta(0) = \theta_0$.

1.6 Deux Lemmes d'Analyse

Nous rappelons simplement ici deux lemmes élémentaires (mais importants) d'analyse qui serons utilisés plus tard.

Lemme .9. ***(Lemme de Césaro)*** *Soient $(u_n)_{n\geq 0}$ une suite convergente et $(a_n)_{n\geq 1}$ une suite de nombres réels positifs croissant vers l'infini. Alors la suite définie par*

$$c_n = \frac{1}{a_n}\sum_{k=1}^{k=n}(a_k - a_{k-1})u_k, \quad n \geq 1,$$

converge vers la même limite que $(u_n)_{n\geq 0}$

Lemme .10. ***(Lemme de Kronecker)*** *Soient $(s_n)_{n\geq 0}$ une suite positive croissant vers l'infini et $(x_n)_{n\geq 0}$ une suite réelle. On suppose que la série de terme général (x_n/s_n) converge. Alors*

$$\frac{1}{s_n}\sum_{k=1}^{k=n} x_k \longrightarrow 0.$$

(voir par exemple [WIL95] p.117)

1.7 Exemple d'Options en finance

Une option est un contrat entre deux ou plusieurs contre-parties portant sur un actif sous-jacent (par exemple une action, un indice, un panier d'indices, une obligation, le cours du pétrole etc...) et qui donne à une des parties une prérogative sur les autres. Les options les plus simples qui existent sont le Call (opiton d'achat) et le Put (option de vente) sur une action.
A priori on peut envisager un nombre infini d'options (l'imagination de l'homme étant sans limite!). Nous nous contenterons ici d'écrire tout simplement le payoff de certaines options dont les options utilisées pour des illustrations numériques dans cette thèse:

- Un Call (*resp.* Put) européen de maturité T est une option sur un sous-jacent S_t, de prix d'exercise (ou strike) K, garantie, à son détenteur, la richesse

  $$\max(S_T - K, 0) \qquad (resp. \quad \max(K - S_T, 0)),$$

  à la date T.

- Un Call asiatique de maturité T et de strike K est un Call standard, mais qui porte sur la moyenne du cours du sous-jacent S entre 0 et T.
  Lorsque la moyenne est continue la richesse finale est donnée par:

$$\psi(S_t,\ t \geq 0) = \left(\frac{1}{T} \int_0^T S_t \, dt - K \right)_+$$

  Alors que qu'elle vaut

$$\psi(S_{t_i},\, i = 1...d) = \left(\frac{1}{d} \sum_{i=1}^{d} S_{t_i} - K \right)_+,$$

  lorsque la moyenne est discrète.

- Un Cap sur un taux d'intérêt r_t, de maturité T, de taux strike K, aux dates de paiement $t_1 < t_2 < ... < t_n$ et $t_{n+1} = T$ sur un nominal de L est une option qui verse à son détenteur le flux $L(r_{t_i} - K)_+$ à chaque date de paiement $t_{i+1}, \quad i = 0, \ldots, n$. Le payoff de ce Cap est la somme de tous les flux actualisés jusqu'à maturité, c'est-à-dire

$$L \sum_{i=1}^{n} \prod_{j=1}^{i} \frac{1}{(1 + r_{t_j})^h} (r_{t_i} - K)_+.$$

- Un Call barrière sur S de maturité T et de strike K donne à son détenteur le droit d'acheter S_T au prix K, à condition que le cours de S soit resté dans un certain domaine D (option OUT) ou en soit sortie (option IN) entre 0 et T. Par exemple, avec une barrière haute (UP) la richesse finale s'écrit:

$$\mathbb{I}_{S_t \leq U} (S_T - K)_+.$$

Toutes ces options (et bien d'autres encore) peuvent être de type américain: dans ce cas l'acheteur peut exercer son droit à toute date comprise entre 0 et T.

Chapitre 2

Fonctions d'Importance et Approximations Stochastiques en Finance

Ce chapitre est l'objet d'un article paru dans la revue "**The Journal of Computational Finance**".

2.1 Introduction

L'usage des simulations Monte Carlo pour l'estimation du prix de certains instruments financiers est apparu en 1977 dans un article de P. Boyle. Aujourd'hui encore les méthodes de Monte Carlo restent très compétitifs dans le pricing et la couverture des produits dérivés en finance, surtout si ces produits sont complexes ou tout simplement dépendent d'un nombre élevé d'actifs sous-jacents. L'efficacité d'une méthode de Monte Carlo étant liée à la variance statistique de l'estimation, l'on essaye, tant que faire se peut, d'utiliser des techniques de réduction de variance pour améliorer sa précision. Le but de ce chapitre, est de présenter l'une de ces techniques. C'est une méthode nouvelle basée sur une formulation simple de la technique des fonctions d'importance et sur l'usage des algorithmes stochastiques de type Robbins-Monro. Les approximations stochastiques sont, depuis longtemps, largement utilisés dans plusieurs domaines des sciences (voir Rubinstein et Melamed [RM98] pour plus de détails), mais leur apparition en finance numérique est récente (1998, Vazquez-Abad et Dufresne [VD98]). Newton (1994 [NEW94]) a montré qu'en théorie dans un problème de pricing en temps continu, la technique des fonctions d'importance peut permettre d'avoir un estimateur de variance nulle par un changement de *drift stochastique*. Toutefois, la détermination de ce paramètre stochastique suppose la connaissance à priori du prix à estimer. Pour mettre

en œuvre cette approche, on utilise donc des approximations (plus ou moins grossières) du prix pour avoir des approximations du drift optimal. L'approche de Glasserman, Heidelberger et Shahabuddin (1999 [GHS99a] et [GHS99b]) est différente: Plutôt que de considérer un changement de paramètre stochastique (en temps continu), ils limitent leurs investigations à un changement de paramètre déterministe fini-dimensionnel. De fait cette restriction ne rend plus possible l'obtention d'estimateurs, du prix de l'option, de variance nulle. Dans le papier que les auteurs ont publié à ce sujet, la méthodologie consiste à exhiber –en utilisant des techniques de Grandes Déviations– une direction d'échantillonnage préférentiel lelong de laquelle une bonne approximation de la vraie variance de l'estimateur Monte Carlo est minimale. Vazquez-Abad et Dufresne (1998 [VD98]) d'une part et Su et Fu (2002 [SF02]) d'autre part sont les premiers à combiner la technique des fonctions d'importance et une approximaton stochastique de type gradient stochastique dans une méthode de Monte Carlo en finance. L'approche que nous présentons ici est semblable à chacune des approches ci-dessus. Tout comme Glasserman, Heidelberger et Shabuddin (1999) nous utilisons la technique des fonctions d'importance dans un cadre (celui du praticien) où la simulation des processus sous-jacents se fait à partir d'une suite de vecteurs gaussiens indépendants. Mais notre méthode de recherche de la direction optimale d'échantillonnage préférentiel est radicalement différente. Elle semble plus proche des travaux de Su et Fu (2002).
L'idée principale de notre approche est d'utiliser une version convenablement adaptée des algorithmes de Robbins-Monro pour optimiser le choix de la direction d'échantillonnage préférentiel. Bien que la vitesse de convergence de ce type d'algorithmes soit en général de l'ordre de $C/\sqrt{n}$, leur implémentation est directe et très facile.
La suite de ce chapitre se décompose comme suit: dans la prochaine section nous présentons brièvement le cadre mathématique et financier suivie, dans la section 3, d'une introduction rapide de la technique des fonctions d'importance (basée sur le théorème de Girsanov). Dans la section 4, nous présentons une méthode de réduction de variance basée sur l'utilisation des algorithmes stochastiques combinée avec la technique des fonctions d'importance. Cette même section contient une preuve du principal résultat théorique de ce chapitre, support de la méthodologie que nous exposons ici.

2.2 Cadres mathématique et financier

Nous modélisons le prix d'un actif sous-jacent directement sous la mesure risque-neutre par l'équation différentielle stochastique suivante

$$dS_t = S_t(rdt + \sigma(t, S_t)dW_t),\ S_0 = x, \tag{.2.1}$$

avec r le taux d'intérêt sans risque, $\sigma(t, y)$ la nappe de volatilité de l'actif, W_t un mouvement brownien standard, et x un réel positif fixé. Pour des raisons d'absence

d'arbitrage, dans un marché sans frictions, le prix d'une option de payoff $\bar{\psi}(S_t, t \leq T)$ est donnée par la formule (voir par exemple [LL96])

$$C_0 = \mathbb{E}[e^{-rT}\bar{\psi}(S_t, t \leq T)], \tag{.2.2}$$

où $\mathbb{E}$ désigne l'espérance prise sous la probabilité risque-neutre.

Remarque .17. L'actif sous-jacent S peut être multidimensionnel. Dans ce cas, chacune de ses coordonnées vérifie une diffusion du type (.2.1) et la formule (.2.2) reste vraie en général.

En pratique, pour évaluer numériquement une telle espérance, il faut le plus souvent discrétiser le processus de sous-jacent $(S_t)_{t\geq 0}$. Nous nous limiterons au cas où la simulation du processus de sous-jacent est simplement gouvernée par une suite de copies indépendantes de la loi gaussienne centrée réduite. Ce n'est d'ailleurs pas un cas particulier puisqu'on peut toujours, par une transformation linéaire, ramener le cas correlé à ce cas. Néanmoins, nous devons garder à l'esprit que lorsqu'une solution de l'équation différentielle stochastique (.2.1) n'est pas connue, la simulation de la diffusion $(S_t)_{t\geq 0}$ se fait après une discrétisation par un schéma numérique de type schéma d'*Euler*. Nous admettons donc qu'une telle discrétisation existe sur une grille de temps $0 = T_0 < T_1 < \cdots < T_d = T$, qu'elle est acceptable, et nous nous focalisons sur l'obtention d'estimateurs précis du prix C_0. Ainsi nous approximons C_0 par $\hat{C}_0$ avec

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

que nous pouvons réecrire comme

$$\hat{C}_0 = \mathbb{E}[\psi(X)], \tag{.2.3}$$

où $X = (X_1, \ldots, X_d) \sim \mathcal{N}(0, I_d)$ et I_d est la matrice identité de $\mathbb{R}^d$. La fonction ψ peut être explicitée à partir de la fonction de payoff $\hat{\psi}$ et de la discrétisation de S.
Nous nous intéressons maintenant à l'évaluation Monte Carlo de (.2.3) par une technique de fonctions d'importance paramétrique.

2.3 Fonction d'Importance

La technique des fonctions d'importance paramétrique consiste à "shifter" la loi de X en lui ajoutant un vecteur $\theta = (\theta_1, \ldots, \theta_d)$ de $\mathbb{R}^d$. Par un simple changement de variable sur $\mathbb{R}^d$ (ou une application d'une version élémentaire du théorème de Girsanov), on peut réecrire (.2.3) sous la forme

$$\hat{C}_0 = \mathbb{E}[g(\theta, X)], \tag{.2.4}$$

avec

$$g(\theta, X) = \psi(X+\theta)e^{-\theta\cdot X - \frac{1}{2}\|\theta\|^2}, \tag{.2.5}$$

$\|x\|$ et $x \cdot y$ désignant respectivement la norme Euclidienne d'un vecteur $x \in \mathbb{R}^d$ et le produit scalaire usuel de deux vecteurs $x, y \in \mathbb{R}^d$. La famille $(g(\theta, X))_{\theta\in\mathbb{R}^d}$ est ainsi une famille paramétrique d'estimateurs sans biais de $\hat{C}_0$. L'idée naturelle est d'essayer d'exhiber d'une telle famille l'estimateur le plus éfficace *i.e.* celui à la variance la plus petite. Dans leur article [GHS99a] Glasserman Heidelberger et Shahabuddin développent une procédure de minimisation de la variance de $g(\theta, X)$ en fonction de θ. L'idée est de chercher à éliminer la contribution de la partie linéaire de la fonction $\ln(\psi(\cdot))$ à la formation de la variance totale de l'estimation. Par des techniques de Grandes Déviations ils montrent que cela revient à faire un échantillonnage lelong de la direction $\hat{\theta}$ de $\mathbb{R}^d$ solution du problème de recherche de point fixe:

$$\nabla \ln \psi(\theta) = \theta, \quad \theta \in \mathbb{R}^d \text{ t.q.} \psi(\theta) > 0.$$

Cette approche consiste en réalité à remplacer le vrai problème de minimisation stochastique par un problème approché de minimisation déterministe. La méthode marche bien dans le cas de fonctions de payoff ψ assez régulières et pour lesquelles la partie linéaire de $\ln(\psi(\cdot))$ est prépondérante.
Notre premier objectif est d'étudier directement le vrai problème de minimisation par des méthodes stochastiques. Si on note dans la suite H la fonction donnée par

$$H(\theta) = \mathbb{E}[g^2(\theta, X)],$$

alors le problème à résoudre est le suivant

$$\min_{\theta\in\mathbb{R}^d} H(\theta). \tag{.2.6}$$

Le résultat qui suit nous permet de convertir ce problème de minimsation en un problème de recherche de zéro d'une fonction.

Proposition 2.3.1. *Supposons que* $\mathbb{P}(\psi(X) > 0) \neq 0$. *Si* $\mathbb{E}(\psi^{2a}(X)) < \infty$, *avec* $a > 1$, *alors* H *est strictement convexe sur* $\mathbb{R}^d$ *et il existe un unique* $\theta^* \in \mathbb{R}^d$ *tel que*

$$H(\theta^*) = \min_{\theta\in\mathbb{R}^d} H(\theta). \tag{.2.7}$$

Remarque .18. En écrivant les conditions d'optimalité de premier ordre pour H on s'aperçoit que le paramètre optimal θ^* est aussi l'unique solution de l'équation

$$\nabla H(\theta) = 0, \quad \theta \in \mathbb{R}^d, \tag{.2.8}$$

Preuve. Un simple changement de variable montre que

$$\begin{aligned} H(\theta) =& \mathbb{E}\left[\psi^2(X+\theta)e^{-2\theta\cdot X-\|\theta\|^2}\right] \\ =& \mathbb{E}\left[\psi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}\right]. \end{aligned}$$

Considérons $\theta \in \mathbb{R}^d$ tel que $\|\theta\| \leq K$ avec $K > 0$. Si on pose

$$F(\theta, x) \ = (\theta - x)\psi^2(x)e^{-\theta\cdot x+\frac{1}{2}\|\theta\|^2},$$

alors on obtient sans difficulté l'inégalité suivante

$$\int |F(\theta,x)|e^{-\frac{1}{2}\|x\|^2}dx \leq e^{\frac{K^2}{2}} \int (K+\|x\|)e^{K\|x\|}e^{-(1-\frac{1}{a})\frac{1}{2}\|x\|^2}\psi^2(x)e^{-\frac{1}{2a}\|x\|^2}dx.$$

Maintenant d'après l'inégalité de Hölder, on a

$$\begin{aligned} \int |F(\theta,x)|e^{-\frac{1}{2}\|x\|^2}dx \leq e^{\frac{K^2}{2}} \left(\int (K+\|x\|)^{\frac{a}{a-1}} e^{\frac{aK}{a-1}\|x\|}e^{-\frac{1}{2}\|x\|^2}dx\right)^{1-\frac{1}{a}} \times \\ \times \left(\int \psi^{2a}(x)e^{-\frac{1}{2}\|x\|^2}dx\right)^{\frac{1}{a}}. \end{aligned}$$

Mais $\mathbb{E}(\psi^{2a}(X)) < \infty$, entraine que la quantité $\int |F(\theta,x)|e^{-\frac{1}{2}\|x\|^2}dx$ est bornée dès que $\|\theta\| \leq K$. D'après le théorème de Lebesgue, on en déduit que la fonction H est différentiable et que

$$\nabla H(\theta) = \mathbb{E}[F(\theta, X)] = \mathbb{E}\left[(\theta - X)\psi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}\right].$$

En réitérant cette argumentation, on montre sans difficulté particulière que H est deux fois continûement différentiable et que sa matrice hessienne est donnée par

$$\text{Hess}H(\theta) = \mathbb{E}\left[\left(I_d + (\theta - X)(\theta - X)^T\right)\psi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}\right],$$

I_d étant la matrice identité de $\mathbb{R}^d$. Comme $\mathbb{P}(\psi(X) > 0) \neq 0$ on a $\forall\ u \in \mathbb{R}^d - \{0\}$, u^T désignant le vecteur transposé de u,

$$u^T \cdot \text{Hess}H(\theta) \cdot u = \mathbb{E}\left[\left(\|u\|^2 + (u\cdot(\theta - X))^2\right)\psi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}\right] > 0,$$

et la matrice hessienne de H est définie positive. Ainsi la fonction H est strictement convexe sur $\mathbb{R}^d$. Pour terminer la preuve il suffit de montrer que

$$\lim_{\|\theta\|\to+\infty} H(\theta) = +\infty.$$

Mais on a

$$\begin{aligned} H(\theta) &\geq e^{\frac{1}{2}\|\theta\|^2}\mathbb{E}\left(\psi^2(X)e^{-\|\theta\|\|X\|}1_{\{\|X\|<\frac{\|\theta\|}{4}\}}\right) \\ &\geq e^{\frac{1}{4}\|\theta\|^2}\mathbb{E}[\psi^2(X)1_{\{\|X\|<\frac{\|\theta\|}{4}\}}]. \end{aligned}$$

Et toujours parce que $\mathbb{P}(\psi(X) > 0) \neq 0$, on voit que $\lim_{\|\theta\|\to+\infty} H(\theta) = +\infty$.
□
On a aussi le corollaire suivant

Corollaire .11. *Si $h(\cdot)$ désigne la fonction définie par*

$$h(\theta) = \nabla H(\theta), \quad \theta \in \mathbb{R}^d, \tag{.2.9}$$

alors on a

$$h(\theta^*) = 0 \quad et \quad \forall\, \theta \in \mathbb{R}^d,\ \theta^* \neq \theta \quad h(\theta)\cdot(\theta - \theta^*) > 0,$$

et la condition de Lyapounov (H_1) du Théorème .5 est vérifiée.

Preuve. C'est une application directe de la formule de la moyenne et de la Proposition 2.3.1.

Remarque .19. Plutôt que de résoudre le vrai problème (.2.7), l'on peut très bien chercher à résoudre le problème approché suivant

$$\min_{\theta\in\mathbb{R}^d} \hat{H}(\theta), \quad \text{avec} \quad \hat{H}(\theta) := \frac{1}{N}\sum_{i=1}^{N} F(\theta, X^{(i)}),$$

où $N \in \mathbb{N}$ est très grand et $X^{(1)}, ..., X^{(N)}$ est un échantillon de copies indépendantes de X. En effet une fois ces trajectoires $X^{(i)}$ générées, le problème approché devient un problème simple d'optimisation déterministe qu'on peut résoudre par des méthodes bien connues d'Analyse numérique. Cette approche est particulièrement efficace lorsque la loi de X est à support *fini* ou même *dénombrable*. Pour plus de précision là dessus l'on pourra se reporter à [SH00] et aux références qui s'y trouvent.

La principale idée de cette partie de notre travail de thèse consiste à exhiber une bonne version de l'algorithme de Robbins-Monro pour évaluer la solution du problème (.2.8), puis à mettre en œuvre une méthode efficace de réduction de variance Monte Carlo.

2.4 Une procédure de Réduction de Variance

D'après la Proposition (2.3.1) et la Remarque (.18), il existe un unique $\theta^* \in \mathbb{R}^d$ tel que $h(\theta^*) = 0$. Ce paramètre nous fournit également l'estimateur Monte Carlo $g(\theta^*, X)$ de variance minimum de la quantité (.2.3), parmi tous les estimateurs de même forme.
En général, la fonction h est numériquement inaccessible même dans les cas les plus simples. On a vu dans le chapitre précédent qu'un algorithme stochastique de la forme (.1.2) peut, dans de tels cas, être utilisé pour évaluer asymptôtiquement le paramètre θ^*.
D'après le Corollaire .11, l'hypothèse (H_1) du Théorème .5 est vérifiée. De plus il est évident que (H_2) est vérifiée pour une large gamme de pas γ_n. Si l'on implémente directement cet algorithme, l'on s'aperçoit parfois qu'il diverge très rapidement. Cela est dû, comme dans l'exemple de la section 1.5.3 au fait qu'une hypothèse de la forme (H_3) ou (H_4) est invérifiable dans ce cas. En effet

$$Y_{n+1} := F(\theta_n, X_{n+1}) = (\theta_n - X_{n+1})\psi^2(X_{n+1})e^{-\theta_n \cdot X_{n+1} + \frac{1}{2}\|\theta_n\|^2}, \qquad (.2.10)$$

est l'*observation de l'algorithme*. Dans cette expression $(\theta_n)_{n\geq 0}$ représente l'agorithme stochastique (.1.2) et $(X_n)_{n\geq 0}$ est une suite *i.i.d.* de vecteurs gaussiens selon la loi de X. On rappelle que si

$$\mathcal{F}_n = \sigma\{\theta_0,\ X_k,\ k \leq n\},$$

il vient facilement que

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

On a observé numériquement que la suite $\left((1 + \|\theta_n\|^2)^{-1}\mathbb{E}[\|Y_{n+1}\|^2/\mathcal{F}_n]\right)_{n\geq 0}$ explose très souvent avec n. Ce qu'on peut expliquer par la forme *"en exponentielle du carré de la norme de θ_n"* de la suite des *observations de l'algorithme*. Pour remédier à cela, nous allons utiliser la méthode de projection de Chen introduite au chapitre précédent.

Remarque .20. On peut obtenir un théorème limite central pour l'algorithme stochastique (.1.2). Une étude théorique plus fine de sa vitesse de convergence montre que la vitesse optimale ($\sim 1/\sqrt{n}$) est atteinte pour des pas de convergence décroissant vers zéro comme $1/n$. Toutefois, en pratique l'on utilise des pas de la forme

$$\gamma_n = \frac{\alpha}{\beta + n} \quad \alpha, \beta > 0.$$

α et β étant des paramètres à régler pour avoir une bonne convergence numérique de l'algorithme. Dans tout ce chapitre et jusqu'au chapitre 3, nous utiliserons essentiellement des pas de convergence de ce type.

Le théorème suivant est le principal résultat de ce chapitre. Il nous permet de mettre en œuvre notre technique de réduction de variance.

Théorème .12. *Supposons que* $\mathbb{P}(\varphi(X) > 0) > 0$. *Si* $\mathbb{E}[|\varphi(X)|^{4a}] < +\infty$ *pour un certain* $a > 1$, *alors en choisissant la suite* U_n *(de la méthode de projection de Chen) à croissance suffisamment lente , l'algorithme défini par (.1.11-.1.13) converge presque sûrement vers l'unique solution* θ^* *de l'équation* $h(\theta) = 0 \quad \theta \in \mathbb{R}^d$.

Remarque .21. Notons que contrairement au résultat ci-dessus, dans le Théorème .4 de [CGG88], la suite U_n peut être n'importe quelle suite croissant vers l'infini. Le choix convenable d'une telle suite apparaîtra tout naturellement dans la preuve que nous allons donner du Théorème .12.

Mais avant la preuve nous pouvons résumer la méthode comme suit: D'abord générer un grand nombre de trajectoires Monte Carlo $(X_n)_{n\geq 0}$

(i) ensuite lancer l'algorithme $(\theta_n)_{n\geq 0}$ defini par (.1.11)-(.1.13) lelong d'un certain nombre de ces trajectoires pour obtenir un estimé $\hat{\theta}$ de θ^*,

(ii) puis utiliser la valeur $\hat{\theta}$ obtenue au (i) dans l'estimation Monte Carlo classique de la quantité recherchée $\hat{C}_0 \simeq \frac{1}{N}\sum_{n=1}^{N} g(\hat{\theta}, X_n)$.

Maintenant voici une preuve du théorème .12

Preuve. Soit $(X_n)_{n\geq 0}$ la suite *i.i.d.* de vecteurs gaussiens centrés réduits utilisée pour générer l'algorithme stochastique $(\theta_n)_{n\geq 0}$. On rappelle que si

$$\mathcal{F}_n = \sigma\{\theta_0,\ X_k,\ k \leq n\},$$

alors θ_n et X_{n+1} sont respectivement $\mathcal{F}_n$-mesurable et indépendant de $\mathcal{F}_n$, de sorte que les *observations de l'algorithme* vérifient

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

où

$$s^2(\theta) = \mathbb{E}[\|\theta - X\|^2 \varphi^4(X) e^{-2\theta\cdot X + \|\theta\|^2}].$$

Notons $f(\theta, x)$ la fonction définie par $f(\theta, x) = \|\theta - x\|^2 e^{-2\theta\cdot x + \|\theta\|^2}$. $\forall \beta > 1$ on a sans difficultés particulières

$$\begin{aligned}
\mathbb{E}(f^\beta(\theta, X)) &= \mathbb{E}(\|\theta - X\|^{2\beta} e^{-2\beta\theta\cdot X + \beta\|\theta\|^2}) \\
&= \mathbb{E}(\|X\|^{2\beta} e^{(2\beta+1)\theta\cdot X - (\beta+\frac{1}{2})\|\theta\|^2}) \quad \text{par un changement de variables} \\
&\leq e^{-(\beta+\frac{1}{2})\|\theta\|^2} \mathbb{E}(\|X\|^{2\beta} e^{(2\beta+1)\|\theta\|\|X\|}).
\end{aligned}$$

Or la variable aléatoire $\|X\|^2$ suit une loi de Khi-deux à d degrés de liberté. Par suite en utilisant l'inégalité $(2\beta+1)\|\theta\|\|X\| \leq (2\beta+1)^2\|\theta\|^2 + \frac{1}{4}\|X\|^2$ on obtient en utilisant la densité de cette loi

$$\begin{aligned}\mathbb{E}(f^\beta(\theta,X)) &\leq C_1(d)e^{-(\beta+\frac{1}{2})\|\theta\|^2+(2\beta+1)^2\|\theta\|^2}\int_0^{+\infty} r^{\beta+\frac{d}{2}-1}e^{-\frac{r}{4}}dr \\ &= C_2(\beta,d)e^{-(\beta+\frac{1}{2})\|\theta\|^2+(2\beta+1)^2\|\theta\|^2}.\end{aligned}$$

D'après l'inégalité de Hölder, il vient pour $p > \frac{1}{4}$

$$s^2(\theta) \leq \left(\mathbb{E}(|\varphi(X)|^{4p})\right)^{\frac{1}{4p}}\left(\mathbb{E}[f^{\frac{4p}{4p-1}}(\theta,X)]\right)^{1-\frac{1}{4p}} \tag{.2.11}$$

$$\leq C(p,d)e^{Q(p)\|\theta\|^2}, \tag{.2.12}$$

avec $C_1(d)$, $C_2(\beta,d)$, et $C(p,d)$ trois constantes strictement positives. $Q(p)$ est une fraction rationnelle (en p) définie par

$$Q(p) := \frac{7.5p^2 - p + 1/32}{p^2 - p/4}.$$

Maintenant en utilisant la définition même de l'algorithme $(\theta_n)_{n\geq 0}$ donnée en (.1.11)-(.1.13) et la condition $U_0 \geq \max(\|\theta^1\|, \|\theta^2\|)$, on voit que

$$\|X_n\| \leq \max(U_{\sigma(n-1)}, \|\bar{\theta}_{n-1}\|) \leq U_n, \quad \text{et} \quad s^2(\theta_n) \leq Ce^{Q(p)U_n^2},$$

pour n assez grand.
Pour finir la preuve, on pose $\epsilon_{n+1} = h(\theta_n) - Y_{n+1}, \quad n \geq 0$ et on définit

$$M_n \;=\; \sum_{i=0}^{n-1}\gamma_{i+1}\epsilon_{i+1} \quad n \geq 1 \quad \text{et} \quad M_0 = 0.$$

Il est clair que $(M_n)_{n\geq 1}$ est une $\mathcal{F}_n$-martingale de carré intégrable. Son processus crochet (oblique) est donné par

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

En exploitant l'estimation (.2.11) on voit qu'on peut choisir la suite (U_n) de sorte que

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

où $C > 0$ est une constante. On conclut, d'après le Théorème .4, que la martingale $(M_n)_{n\geq 0}$ converge presque sûrement, *i.e.*,

$$\lim_{n\to+\infty}\ \sum_{i=0}^{n-1}\gamma_{i+1}\epsilon_{i+1} < +\infty, \quad p.s.$$

Le Lemme .10 de Kronecker entraine alors que

$$\lim_{n\to+\infty}\gamma_{n+1}\left\|\sum_{i=0}^{n-1}\epsilon_{i+1}\right\| = 0 \quad p.s.$$

Ainsi l'hypothèse (H_6) est vérifiée et le Théorème .6 s'applique. □

Remarque .22. Le choix de la suite U_n de projection se fait en étudiant le comportement de la fraction rationnelle $Q(p) = \frac{7.5p^2-p+1/32}{p^2-0.25p}$. En effet on peut montrer que

$$\forall\, p > 1, \quad Q(p) \in]7.5,\ 8.71[,$$

De sorte que le choix $U_n = \sqrt{\frac{1}{10}\ln n} + M + 10,\ n \geq 1$ entraine

$$\forall\, p > 1,\ e^{Q(p)U_n^2} < e^{9U_n^2} = Constant \times n^{\frac{9}{10}},\ \text{et la série}\ \sum_n \gamma_n^2 e^{Q(p)U_n^2}\ \text{converge},$$

puisque $\gamma_n^2 e^{Q(p)U_n^2} \sim (1/n)^{\frac{11}{10}}$ au voisinage de l'infini.

2.5 Tests numériques

A travers nos illustrations numériques nous voulons voir si l'apport de la méthode de réduction de variance que nous proposons ici est significatif. Nous mesurons cet apport en terme de rapport de variances de notre méthode et de la méthode de Monte Carlo standard. Ce cœfficient est noté **RV** dans le reste de ce chapitre (et même dans toute

cette thèse). Nous verrons qu'il est suffisant pour la comparaison. Le cœfficient **RV** est défini par

$$\mathbf{RV} := \frac{\frac{1}{N}\sum_{n=1}^{n=N} g^2(0, X_n) - \bar{C}_N^2}{\frac{1}{N}\sum_{n=1}^{n=N} g^2(\bar{\theta}, X_n) - \bar{C}_N^2},$$

avec $\bar{\theta}$ une estimation asymptôtique de θ^*, $g^2(\theta, x) := \varphi^2(x)e^{-\theta\cdot x+\frac{1}{2}\|\theta\|^2}$, et $\bar{C}_N = \frac{1}{N}\sum_{n=1}^{n=N}\varphi(X_n)$. Evidemment **RV** est une approximation du vrai coefficient $\mathbf{RV}^*$

$$\mathbf{RV}^* := \frac{\mathbb{V}\mathrm{ar}[g(0,X)]}{\mathbb{V}\mathrm{ar}[g(\theta^*,X)]} = \frac{H(0) - \hat{C}_0^2}{H(\theta^*) - \hat{C}_0^2},$$

lequel ne peut manifestement pas être calculé explicitement. On peut toutefois montrer que

$$|\mathbf{RV} - \mathbf{RV}^*| \xrightarrow[n\to+\infty]{p.s.} 0.$$

L'implémentation numérique de l'algorithme (.1.11-.1.13) requiert une attention particulière sur le choix de la suite de pas de convergence $(\gamma_n)_{n\geq 0}$. Ce choix reste toujours délicat. D'un point de vue théorique, la question est réglée: on montre que la meilleure suite (celle qui permet d'avoir une vitesse optimale de convergence en loi dans le TCL de l'algorithme) doit décroitre vers 0 comme $1/n$. Mais comme nous l'avons vu à la Remarque .20, dans les applications, il est plus commode de choisir des pas du type

$$\gamma_n = \frac{\alpha}{\beta + n}, \quad \alpha, \beta > 0,$$

α et β étant des paramètres à régler pour obtenir une bonne convergence numérique de l'algorithme. Dans nos tests, nous avons observé que des variations du paramètre β affectait peu cette convergence, contrairement à α. La question générale du choix empirique du bon α reste un problème ouvert sur lequel nous reviendrons un peu à la fin du prochain chapitre. Ce choix est lié à une bonne connaissance des ordres de grandeur des variables (des observations) qui interviennent dans la mise en œuvre de l'algorithme. Nous nous limitons, dans ce qui suit, à donner quelques indications utiles sur ce choix en pricing d'options.
Notons d'abord que nous avons observé que les valeurs de α dépendent plus du ratio S_0/K que des autres paramètres de l'options, K étant le prix auquel l'option est exercée. Par suite, les valeurs du paramètre α que nous donnons ici peuvent être utilisées pour le pricing Monte Carlo des options de même nature que celle pour lesquelles ces valeurs ont été trouvées à condition que les ratio S_0/K soient du même ordre de grandeur que ceux utilisés ici. L'on pourra ainsi pousser les simulations plus loin en tabulant les valeurs de α correspondant au ratio S_0/K.

Dans tous nos tests, les points de réinitialisation (.1.13) de l'algorithme de Chen que nous avons utilisés sont

$$x^1 = 0.00925\mathbf{e} \quad \text{et} \quad x^2 = 0.0725\mathbf{e},$$

pour les call et

$$x^1 = -0.00925\mathbf{e} \quad \text{et} \quad x^2 = -0.0725\mathbf{e},$$

pour les put. $\mathbf{e}$ désigne le vecteur $(1, ..., 1)$ de $\mathbb{R}^d$. Nous rappelons que le choix de ces vecteurs est complètement arbitraire. Mais une fois ces vecteurs fixés, les valeurs de la constante M utilisées (voir (.1.10)) sont $M = 50$ et $M = 100$ respectivement pour les modèles à volatilité constante et à volatilité stochastique.

Remarque .23. Rappelons que l'objectif de notre méthode est de minimiser la fonction $H(\theta) = \mathbb{E}[g^2(\theta, X)]$ par un algorithme stochastique de type Robbins-Monro de la forme

$$\theta_{n+1} = \theta_n - \frac{\alpha}{n+1} F(\theta_n, X_{n+1}),$$

où $\alpha > 0$ et $F(\theta, x)$ est telle que $\nabla H(\theta) = \mathbb{E}[F(\theta, X)]$. Il est clair que les itérations ci-dessus ont une variabilité importante selon que H est susceptible de prendre ou non de grandes valeurs; ce qui dégrade la convergence numérique de l'algorithme. Une façon de réduire ce fait peut consister à appliquer à H (et donc à F) une homothétie de rapport strictement plus petit que 1. Par exemple en pricing d'options, pour des call ou put, si nous posons $\bar{\varphi}(x) = \varphi(x)/S_0$, S_0 étant la valeur spot du sous-jacent, alors

$$H(\theta) = S_0^2 \bar{H}(\theta),$$

avec

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

et

$$\underset{\theta}{\text{argmin}} H(\theta) = \underset{\theta}{\text{argmin}} \bar{H}(\theta).$$

La remarque consiste alors à implémenter l'algorithme (.1.11-.1.13) pour estimer θ^* en utilisant les observations issues de $\bar{H}$ plutôt que de H.
Ainsi pour avoir les "vraies" valeurs du paramètre α de la suite de pas de convergence, il faut diviser les valeurs qui apparaissent sur l'axe des abscisses de toutes les figures par S_0^2.

Remarque .24. Dans toutes nos illustrations numériques, la valeur initiale de l'algorithme (.1.11-.1.13) utilisée est $\theta_0 = 0$. Néanmoins dans certains cas, il est possible d'avoir une idée intuitive d'un bon "voisinage" du paramètre optimal θ^*. Il est, dans de tels cas, plus naturel de choisir θ_0 dans ce voisinage.

2.5.1 Modèle à volatilité constante

Nous étudions ici la fiabilité et la robustesse de la méthode que nous proposons. Pour des Call et Put Européens et Asiatiques, nous traçons le ratio **RV** en fonction du paramètre α (noté "alpha" sur les figures) de la suite de pas de convergence mais avec des trajectoires simulées différentes pour chaque courbe. L'objectif est de mesurer le degré de stabilité de la réduction de variance obtenue en fonction des tirages Monte Carlo utilisés. Sur la Figure 2.1, apparaîssent les courbes de $\sqrt{\mathbf{RV}}$ obtenues pour un Put européen à la monnaie, tandis que la figure 2.2 représente les mêmes courbes pour un Call européen en dehors de la monnaie. Chacune des courbes sur ces figures sont traçées avec 50 000 trajectoires Monte Carlo dont 10 000 ont servi à faire converger l'algorithme de Robbins-Monro. Les figures 2.3 et 2.4 représentent le $\sqrt{\mathbf{RV}}$ dans le cas multidimensionnel d'un Call et Put asiatiques respectivement. Ces graphiques sont basés sur 1 000 000 de trajectoires simulées au total. On remarque que globalement, la réduction de variance obtenue est bonne et effective pour une large fourchette de valeurs de α. La figure 2.5 qui suit représente le $\sqrt{\mathbf{RV}}$ traçé avec les mêmes tirages Monte Carlo dans le cas d' un Call européen à la monnaie sur un panier de 10 actifs sous-jacents. Nous avons utilisé un total de 90 000 tirages Monte Carlo dont respectivement 2 500, 5 000, et 10 000 tirages ont servi pour l'estimation asymptôtique du drift optimal θ^*. Il apparait clairement sur cet exemple que le taux de réduction de variance qu'on obtient diminue avec l'augmentation de la valeur du paramètre α (de la suite de pas de convergence de l'algorithme de Robbins-Monro) et la baisse du nombre d'itérations effectuées pour estimer ce drift optimal.
Cela souligne une variabilité importante de l'algorithme stochastique utilisé. Une façon de réduire cette variabilité est de trouver une observation aléatoire $\hat{Y}_{n+1}$ telle que

$$\mathbb{E}[\hat{Y}_{n+1}/\mathcal{F}_n] = h(\theta_n) \quad \text{et} \quad \mathbb{E}[\hat{Y}^2_{n+1}/\mathcal{F}_n] < \mathbb{E}[Y^2_{n+1}/\mathcal{F}_n], \tag{.2.13}$$

et de l'utiliser à la place de l'observation initiale Y_{n+1}. En effet, si la suite de variables $(Y_{n+1})_{n\geq 0}$ était déterministe, alors l'algorithme (.1.2) serait simplement un schéma d'Euler à pas décroissant de recherche des racines d'une fonction. Nous reviendrons un peu plus en détail sur l'étude de ce phénomène de variabilité à la fin du Chapitre 2.

Remarque .25. Il est important de noter que sur toutes les figures, nous avons toujours $\sqrt{\mathbf{RV}} \geq 1$.

Figure 2.1: Stabilité du ratio de variance $\sqrt{\mathbf{RV}}$ pour un Put européen à la monnaie. Les paramètres sont $\mathbf{S_0 = strike = 50},\ \mathbf{r = 0.10},\ \mathbf{T = 1},\ \sigma = \mathbf{15}\%$

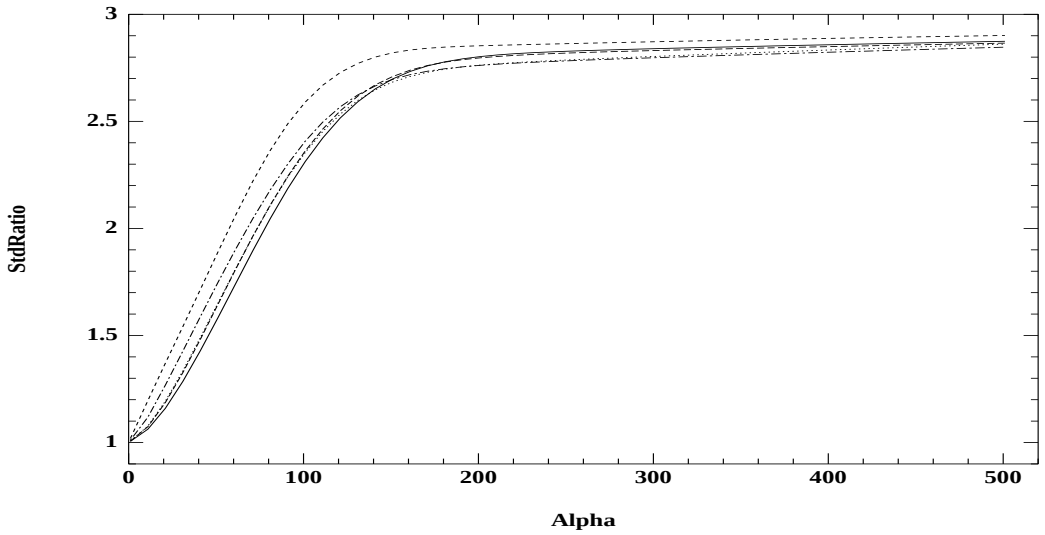


Figure 2.2: Stabilité du ratio de variance $\sqrt{\mathbf{RV}}$ pour un Call européen en dehors de la monnaie. Les paramètres sont $\mathbf{S_0 = 50},\ \mathbf{strike = 60},\ \mathbf{r = 0.05},\ \mathbf{T = 1},\ \sigma = \mathbf{30}\%$.

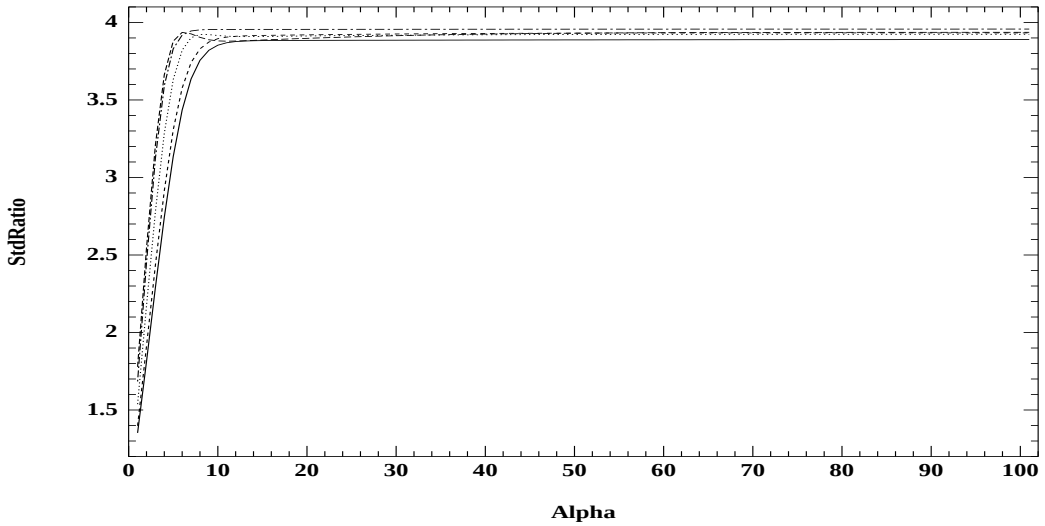


Figure 2.3: Stabilité du rapport de variance **RV** pour un Call asiatique à la monnaie. Les paramètres sont $\mathbf{n = 20},\ \mathbf{S_0 = K = 50},\ \mathbf{r = 0.05},\ \sigma = \mathbf{30}\%,\ \mathbf{T = 0.5}$.

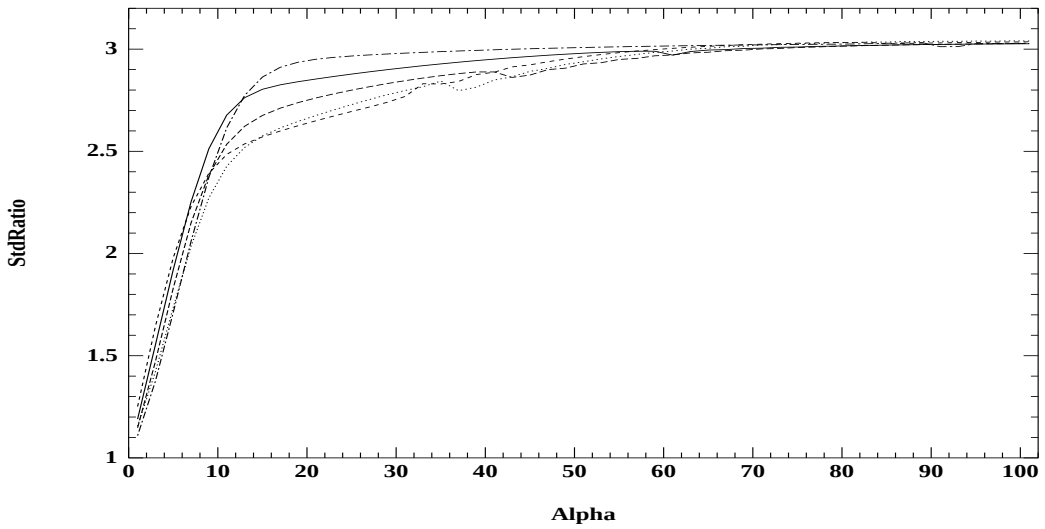


Figure 2.4: Stabilité du ratio de variance **RV** pour un Put asiatique à la monnaie. Les paramètres sont $\mathbf{n = 20},\ \mathbf{S_0 = K = 50},\ \mathbf{r = 0.05},\ \sigma = \mathbf{10}\%;\ \mathbf{T = 0.5}$.

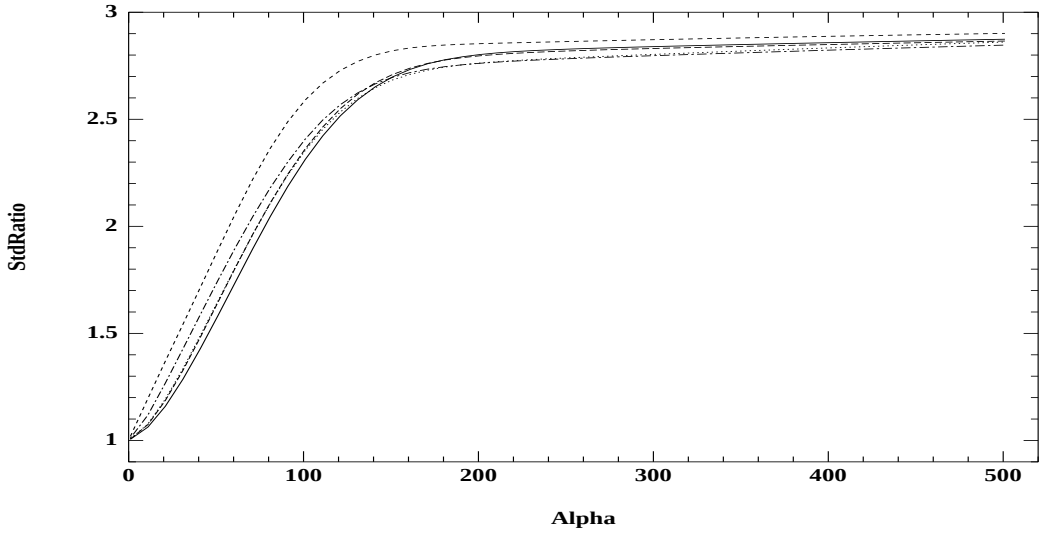
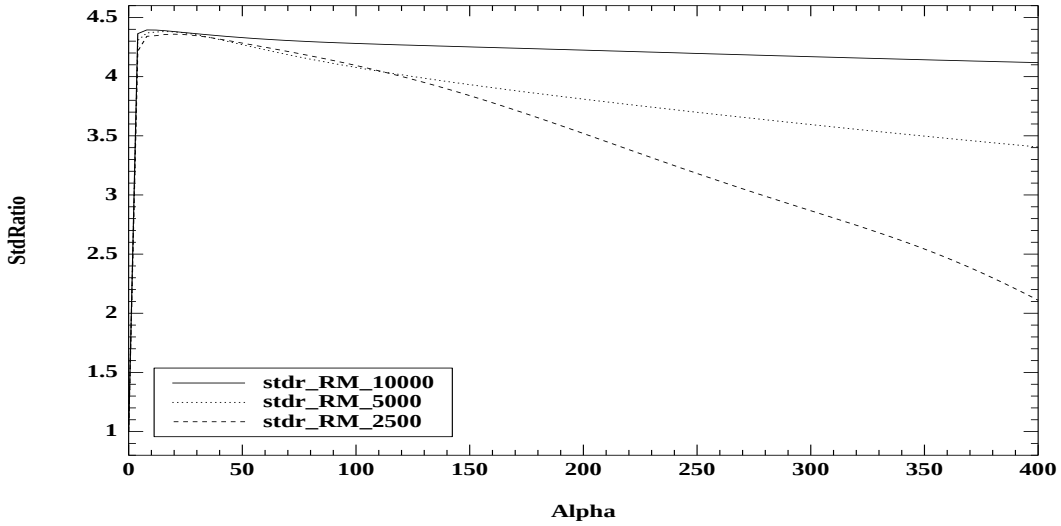


Figure 2.5: Taux de réduction de variance $\sqrt{\mathbf{RV}}$ pour un Call basket à la monnaie par rapport au nombre total de trajectoires Robbins-Monro. Les paramètres sont $\mathbf{K = 50},\ \mathbf{r = 0.05},\ \mathbf{S_0 = 50},\ \mathbf{T = 1}$.



2.5.2 D'autres résultats sur la volatilité constante

Les résultats de la suite de nos investigations numériques se présentent sous forme de tableaux. Les Tableaux 1 et 2 présentent les résultats obtenus pour un Put et un Call européens respectivement. Evidemment les formules de pricing de ces options existent analytiquement, mais il semble naturel de commencer les essais par ces exemples parmi les plus simples afin de mesurer le gain en terme de réduction de variance mais aussi l'absence de biais dans les estimateurs fournis par notre méthode. Dans ces tableaux, **PrixRM**, **PrixMC** et **PrixBS** désignent respectivement l'estimation du prix par notre méthode (Monte Carlo + Fonction d'Importance + Algorithme Stochastique), par la méthode de Monte Carlo standard et le prix Black & Scholes. Nous rappelons que **RV** désigne le rapport des variances statistiques de notre méthode et de celle de la méthode de Monte Carlo naïve. Sur ces exemples, on remarque clairement que ce ratio est très intéressant. De plus nos prix sont nettement plus précis que ceux obtenus à l'aide de la méthode Monte Carlo naïve. le ratio **RV** est d'environ 5 dans le plus mauvais des cas et peut dépasser un facteur de 600 dans les cas les plus favorables (options loins de la monnaie). Le Tableau 3 contient les taux de réduction de variance pour un Call Asiatique (sur moyenne arithmétique). Comme on peut le constater le ratio **RV** est au moins égal à un facteur de 6. Ce qui signifie pour ce cas qu'a un niveau de

Tableau 1 Réduction de Variance estimée pour un Put Européen dans le modèle de Black–Scholes

Paramètres				IS			
α	β	σ	K/S_0	PrixRM	PrixBS	PrixMC	RV
1500	1.	0.3	0.6	0.13	0.13	0.13	38.4
50			0.8	1.28	1.28	1.27	10.9
50			1.0	4.68	4.68	4.70	6.3
10			1.2	10.54	10.53	10.58	4.8
5000		0.1	0.8	0.0042	0.0042	0.0040	350.
50			1.0	0.97	0.96	0.96	9.6
50			1.2	7.31	7.31	7.33	6.3

Les résultats utilisent 40,000 tirages Monte Carlo et 10,000 itérations Robbins-Monro. Les paramètres sont: $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

précision fixé, l'on a besoin approximativement de 6 fois moins de simulations dans notre méthode que pour la méthode de Monte Carlo classique. Dans le Tableau 4 nous exposons les mêmes résultats, mais pour un Call européen sur panier. Dans le cadre du

modèle à volatilité constante, le pricing des options sur panier est typiquement l'un des problèmes d'estimation pour lesquels une méthode de Monte Carlo est inévitable. Pour obtenir des estimations fiables, il faut donc disposer d'une technique de réduction de variance efficace et systématique. La nappe de volatilité est prise "flat" à **15**% ou **30**% et la valeur spot des sous-jacents est identique et égale à $\mathbf{S_0}$. Les résultats contenus dans ce tableau confirment l'intérêt de notre méthode. Dans tous les cas le ratio de variance est supérieur à 10.
La complexité (au sens algorithmique du terme) d'un algorithme stochastique est équivalente à celle d'une méthode de Monte Carlo. Soient **RM** et **MC** respectivement le nombre d'itérations de l'algorithme stochastique et de la méthode de Monte Carlo, si **N** désigne le nombre de pas de discrétisations du modèle, alors la complexité de notre méthode est de l'ordre de $\mathbf{Constant} \times \mathbf{N} \times (\mathbf{RM} + \mathbf{MC})$. Dans tous nos tests, **RM** est au plus égal à 20% de **MC**. Parallèlement le taux de réduction de variance a atteint dans le cas le moins favorable un facteur de 5 et un facteur de quelques centaines dans les cas les plus intéressants.

Tableau 2 Réduction de Variance estimée pour un Call Européen sous le modèle de Black-Scoles

Paramètres				IS			
α	β	σ	K/S_0	PrixRM	PrixBS	PrixMC	RV
25	1.	0.3	0.6	21.63	21.60	21.52	16.8
25			1.0	7.12	7.12	7.01	10.9
50			1.2	3.45	3.45	3.43	15.2
50			1.6	0.67	0.67	0.68	46.2
10		0.1	0.6	21.47	21.46	21.52	112.4
20			1.0	3.41	3.40	3.38	7.8
100			1.2	0.23	0.23	0.23	31.4
5000			1.4	0.004	0.004	0.004	625.0

Les résultats utilisent 40,000 tirages Monte Carlo et 10,000 itérations Robbins-Monro. Les paramètres sont: $\mathbf{S_0 = 50},\ \mathbf{r = 0.05},\ \mathbf{T = 1.0}$.

Tableau 3 Réduction de Variance estimée pour un Call Asiatique arithmétique sous le modèle de Black-Scoles

	Paramètres				IS	
n	α	β	σ	K/S_0	PrixRM	RV
20	25	1.	0.1	0.8	10.723	57.8
	250			1.0	1.53	7.3
	5000			1.1	0.037	6.3
	25		0.3	0.8	10.77	14.4
	250			1.0	2.99	9.0
	2500			1.2	0.320	28.1
40	25		0.1	0.8	10.75	13.0
	250			1.0	2.99	9.0
	2500			1.2	0.321	24.0
	25		0.3	0.8	10.72	27.0
	250			1.0	1.53	6.8
	10000			1.2	0.037	7.8

Les résultats sont basés sur 40,000 tirages Monte Carlo et 20,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{S_0 = 50},\ \mathbf{r = 0.1},\ \mathbf{T = 0.5}$.

Tableau 4 Réduction de Variance estimée pour un Call Basket Européen sous le modèle de Black-Scoles

		Paramètres			IS	
n	α	β	σ	K/S_0	PrixRM	RV
10	2.5	1.	0.3	0.6	28.03	37.2
	2.5			0.8	23.47	34.8
	2.5			1.0	19.85	34.8
	2.5			1.2	16.96	36.0
	2.5			1.4	14.60	37.2
	2.5		0.15	0.6	23.66	16.8
	2.5			0.8	16.74	13.7
	2.5			1.0	11.52	12.3
	25			1.2	7.78	16.8
	25			1.4	5.22	20.3
20	2.5		0.3	0.6	32.24	90.3
	2.5			0.8	28.88	84.6
	2.5			1.0	26.14	84.6
	25			1.2	23.85	67.2
	25			1.4	21.90	70.6
	2.5		0.15	0.6	25.21	21.2
	2.5			0.8	19.46	19.4
	2.5			1.0	16.02	19.4
	25			1.2	11.65	20.1
	25			1.4	9.09	25.0

Les résultats sont basés sur 90,000 tirages Monte Carlo et 10,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

2.5.3 Une comparaison avec la méthode GHS

Rappelons tout d'abord que le problème de départ était le suivant

$$\min_{\theta \in \mathbb{R}^d} \mathbb{E}\left[\varphi^2(X) e^{-\theta \cdot X + \frac{1}{2}\|\theta\|^2} \mathbf{1}_D\right], \tag{VP}$$

où $D = \{x \in \mathbb{R}^d \, \varphi(x) > 0\}$. En posant $\varphi(X)\mathbf{1}_D = \exp(\phi(X))\mathbf{1}_D$, par la formule de Laplace classique (voir Bleistein et Handelsman 1975, ch 8.) on voit que

$$\mathbb{E}\left[\varphi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}\mathbf{1}_D\right] = (2\pi)^{d/2}\int_D e^{2\phi(x)-\theta\cdot x+(1/2)\|\theta\|^2-(1/2)\|x\|^2}dx$$
$$\simeq \mathsf{const}\times\exp\left(\max_{x\in D}\{2\phi(x)-\theta\cdot x+(1/2)\|\theta\|^2-(1/2)\|x\|^2\}\right).$$

Dans [GHS99a] les auteurs remplacent le problème de minimisation stochastique (VP) par le problème approché suivant

$$\min_{\theta\in\mathbb{R}^d}\max_{x\in D}\left\{2\phi(x)-\theta\cdot x+\frac{1}{2}\|\theta\|^2-\frac{1}{2}\|x\|^2\right\}.$$

En suite en utilisant des techniques de grandes déviations, ils développent un context dans lequel le problème min-max ci-dessus est équivalent à

$$\max_{x\in D}\left\{\phi(x)-\frac{1}{2}\|x\|^2\right\}.$$

Ce dernier problème d'optimisation est déterministe et facile à résoudre par des techniques usuelles. Il faut souligner tout de même que cette méthode exige un certain degré de régularité de la fonction de payoff φ. Il est donc clair que notre méthode sera meilleure à celle de [GHS99a] pour le pricing des options de payoffs irréguliers. Ce que nous faisons ici, est simplement de comparer les deux méthodes dans le cas asiatique. Cette comparaison est résumée par le Tableau 5. $\mathbf{IS}+\mathbf{RM}$ et $\mathbf{IS}+\mathbf{GHS}$ correspondent respectivement aux résultats relatifs à notre méthode et à celle de [GHS99a]. On y voit que, dans ce cas particulier les taux de réduction de variance sont équivalents pour les deux méthodes. Nous avons poussé l'étude en comparant les drifts optimaux donnant les directions optimales d'échantillonnage dans les deux méthodes. Les figures 2.6 et 2.7 représentent ces deux directions optimales d'échantillonnage dans le cas des options (Asiatiques) sur moyenne arithmétique. On voit que ces directions sont presque identiques. D'ailleurs le coefficient de corrélation des vecteurs correspondant à ces directions est de $\mathbf{99.97}\%$. Nous avons utilisé un ordinateur Pentium III 1Ghz / 1Go Ram pour faire nos tests et les temps de calcul sont de 1s64 pour la méthode GHS et de 1s92 pour notre méthode.

Table 5 Réduction de Variance estimée pour un Call Asiatique sous le modèle de Black-Scoles. Comparaison entre notre méthode stochastique et la méthode déterministe de [GHS99a]

Paramètres			IS+RM		IS+GHS	
σ	α	K/S_0	Price	RV	Prix	RV
	50	0.9	6.054	11.4	6.052	10.6
0.1	200	1.0	1.919	7.6	1.921	7.0
	2.10^4	1.1	0.203	22.3	0.201	21.1
	20	0.9	7.144	8.9	7.145	8.3
0.3	100	1.0	4.174	10.1	4.169	9.2
	500	1.1	2.211	13.0	2.206	12.0

Les résultats sont basés sur 42,500 tirages Monte Carlo et 20,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{n = 16},\ \mathbf{S_0 = 50},\ \mathbf{T = 1.0},\ \mathbf{r = 0.05}$.

Figure 2.6: Drifts optimaux avec IS+RM (trait continu) et IS+GHS (trait discontinu) pour une option Asiatique. Les paramètres sont $\mathbf{n = 16},\ \mathbf{S_0 = K = 50},\ \mathbf{r = 0.05},\ \sigma = \mathbf{0.1},\ \alpha = \mathbf{200},\ \mathbf{T = 1},\ \mathbf{20,000}$ itérations.

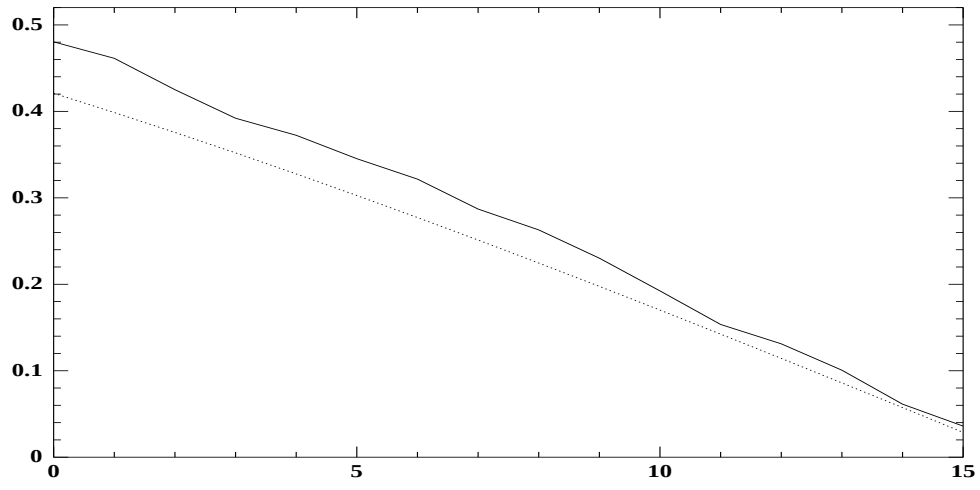
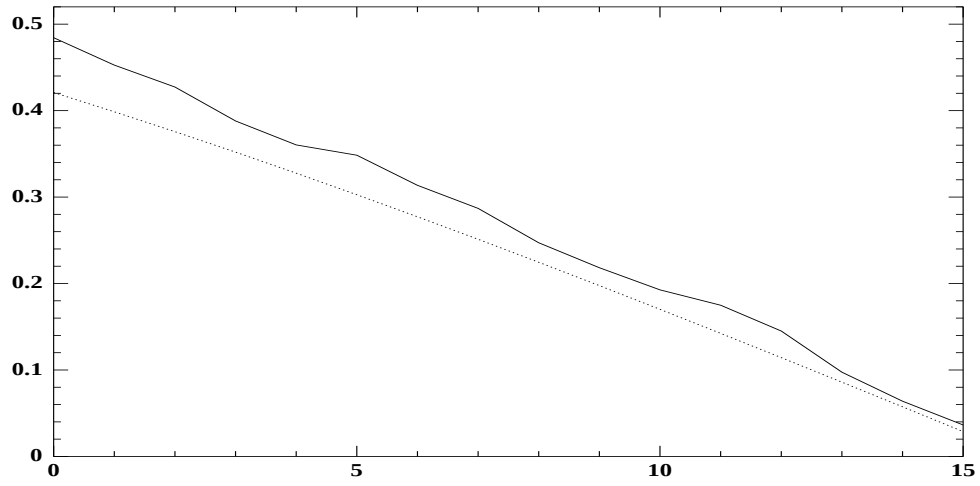


Figure 2.7: Drift optimaux avec IS+RM (trait continu) et IS+GHS (trait discontinu) pour une option Asiatique. Les paramètres sont $\mathbf{n = 16}$, $\mathbf{S_0 = K = 50}$, $\mathbf{r = 0.05}$, $\sigma = \mathbf{0.3}$, $\alpha = \mathbf{100}$, $\mathbf{T = 1}$, $\mathbf{20,000}$ itérations.



2.5.4 Modèles à volatilité stochastique

Le premier exemple considéré, dans ce cadre, est le modèle à volatilité stochastique de Heston (1993). Nous avons testé notre méthode sur un Call européen. Ce choix est motivé par le fait qu'il existe une formule fermée pour ce type d'option sous le modèle de Heston. On constate une fois de plus que notre technique de réduction de variance donne des prix assez précis.

Modèle de Heston

Sous ce modèle l'actif sous-jacent est gouverné par la dynamique suivante

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

où W^1 et W^2 sont deux mouvements browniens correlés avec $\langle W^1, W^2 \rangle_t = \rho t$ et k, $\bar{v}$ et σ des constantes réelles positives. Pour mettre en œuvre notre méthode, nous commençons par discrétiser les diffusions ci-dessus par un schéma d'Euler simple

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

avec $(X^i)_{i\geq 1}$ une suite de copies indépendantes de la loi normale centrée réduite. Dans notre implémentation nous utilisons en entrée, pour chaque itération, un vecteur de taille $2d$ $(X^1, \ldots, X^{2d})$. Le type d'option considéré est un put sur la moyenne arithmétique

$$\hat{S} = \frac{1}{d}\sum_{i=1}^{d} S_{t_i}.$$

Les résultats que nous avons obtenu sont regroupés dans le Tableau 6. Dans ce Tableau, ClosPR désigne le prix obtenu par la formule fermée de Heston. Une fois de plus il apparaît que notre méthode marche bien.
Nous concluons ces tests avec le modèle à volatilité stochastique de Hull et White (1987 [HW87]).

Modèle de Hull et White

Ce modèle est défini par

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

où W^1 et W^2 sont deux mouvements browniens correlés avec $\langle W^1, W^2\rangle_t = \rho t$.
Dans ce modèle, le sous-jacent S est d'espérance finie, mais de variance infinie. Dans un cadre discrèt, on remarque que cette variance est finie mais augmente très vite avec le nombre de pas de discrétisations. Pour atténuer cet effet, on peut tronquer la discrétisation comme suit (voir [GHS99a])

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

avec c une constante réelle positive non nulle. L'impact d'une telle troncature sur la valeur moyenne du sous-jacent est insignifiante, mais elle stabilise les variances estimées. En prenant $\zeta = 0$, on retombe dans le cas de la volatilité constante. Une fois ce travail effectué, l'implémentation de la méthode est presque aussi simple que dans le cas du modèle à volatilité constante. Nous avons, de nouveau, testé la robustesse de la méthode par rapport aux trajectoires Monte Carlo simulées en traçant le ratio $\sqrt{\mathbf{RV}}$ avec différents jeux de ces trajectoires. Dans les simulations effectuées nous avons utilisé les paramètres suivants $c = 2$, $\nu = 0$, $S_0 = 50$, $T = 0.5$, et $\rho = 0.5$.

Table 6 Réduction de Variance estimée pour un Call Européen sous le Modèle à volatilité stochastique de Heston

Paramètres				IS		
$\mathbf{v_0}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\mathbf{K/S_0}$	**ClosPR**	**PrixRM**	**RV**
0.01	10	1.	0.8	11.95	11.94	9.0
	50		0.9	7.27	7.27	7.3
	100		1.0	3.33	3.33	8.4
	500		1.1	1.13	1.12	13.0
	500		1.2	0.32	0.32	16.0
0.04	12.5		0.8	12.00	12.00	9.0
	25		0.9	7.67	7.66	8.4
	50		1.0	4.27	4.25	9.0
	250		1.1	2.11	2.10	11.6
	500		1.2	0.95	0.95	14.4

Le nombre de pas de discrétisations vaut n=100.Les résultats sont basés sur 15,000 tirages Monte Carlo et 5,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{k = 2.0}$, $\theta = \mathbf{0.01}$, $\rho = \mathbf{0.5}$, $\mathbf{T = 1.0}$, $\sigma = \mathbf{0.1}$.

2.6 Remarques et premières conclusions

L'originalité de la méthode de réduction de variance que nous proposons dans ce chapitre provient de l'utilisation d'une bonne version des algorithmes de Robbins-Monro. Nous avons prouvé théoriquement et numériquement que la méthode marche. Nous verrons au Chapitre 3 qu'elle pourrait être améliorée, par un travail préalable sur la version des algorithmes stochastiques utilisée. Nous verrons également qu'en combinant notre méthode avec une méthode de réduction de variance standard, le taux de réduction obtenu peut être très grand. On peut enfin observer que cette méthode réduit le plus souvent la variance de l'estimation Monte Carlo d'une espérance $\mathbb{E}[\varphi(X)]$; en tout cas elle ne l'augmente pas. (Lorsque la fonction $\varphi(\cdot)$ est symétrique, il est facile de voir que la solution du problème (.2.7) n'est autre que $\theta^* = 0$. Cela signifie tout simplement qu'il vaut mieux ne pas appliquer de technique de fonction d'importance pour la réduction de variance dans ce cas! Puisque la probabilité initiale sous laquelle les simulations sont effectuées est déjà en ce sens, une probabilité optimale.

Figure 2.8: Stabilité du ratio de Variance **RV** pour un Put Asiatique en dehors de la monnaie sous le modèle de Hull–White à $\zeta = \mathbf{50}\%$. Les autres paramètres sont $\sqrt{\sigma_0} = \mathbf{0.4}$, $\mathbf{r = 0.1}$.

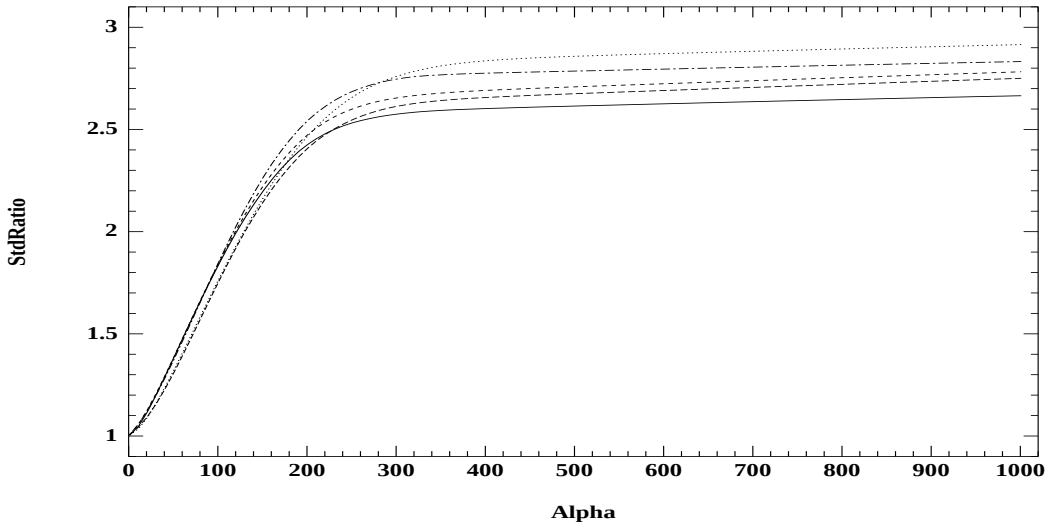
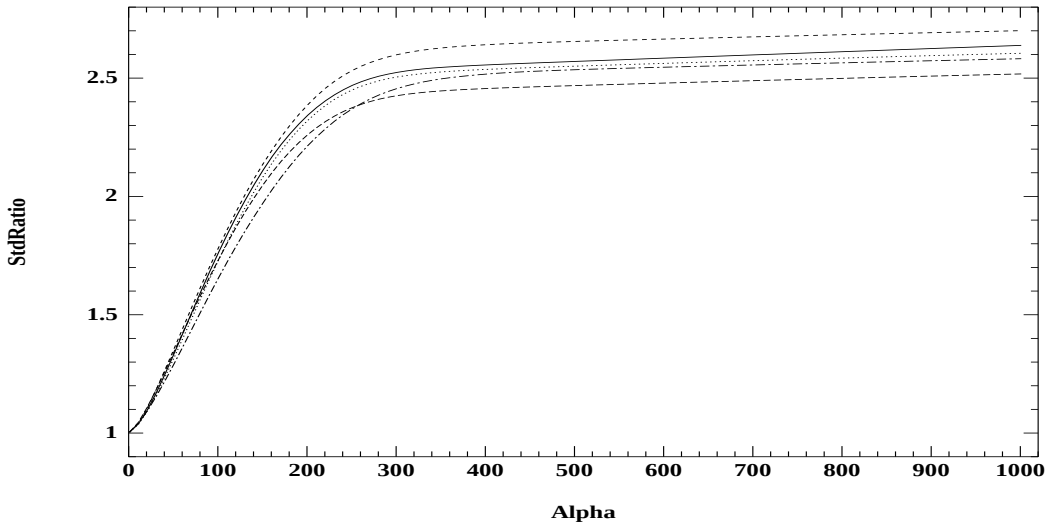


Figure 2.9: Stabilité du ratio de Variance **RV** pour un Put Asiatique en dehors de la monnaie sous le modèle de Hull–White à $\zeta = \mathbf{100}\%$. Les autres paramètres sont $\sqrt{\sigma_0} = \mathbf{0.4}$, $\mathbf{r = 0.1}$.



Chapitre 3

Méthode de Monte Carlo Adaptative

Ce chapitre a fait l'objet d'un article paru dans la revue "**Monte Carlo Methods and Applications**".

3.1 Introduction

Dans ce chapitre nous présentons une version adaptative de la méthode de réduction de variance exposée dans le chapitre précédent. Cette nouvelle méthode est justifiée par des résultats théoriques généraux.
Le chapitre se décompose comme suit. Dans la section 2 nous exposons les motivations qui nous ont poussé à développer une telle méthode. La section 3 contient les résultats principaux du chapitre et leurs démonstrations. Dans la section 4 nous expliquons comment ces résultats permettent de proposer une technique de réduction de variance. Nous illustrons cette méthode par des applications en finance et en fiabilité. Enfin la dernière section contient nos conclusions, des remarques pratiques et une étude de la variabilité de l'algorithme stochastique utilisé.

3.2 Motivations

Considérons le problème général de l'estimation d'une quantité de la forme

$$\mathbb{E}[\varphi(X)],$$

où X suit une loi quelconque simulable et supposons qu'il existe une transformation paramétrique de la forme

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

Remarque .26. En pratique une telle représentation est assez facile à obtenir dès que la loi de X possède une densité.

1. Par exemple si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ avec une densité multivariée $f(x) > 0$, alors on a simplement

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

   avec $g(\theta, x) = \varphi(x + \theta)\frac{f(x+\theta)}{f(x)}$.

2. Un autre exemple consiste à considérer la transformation matricielle suivante

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

   avec $g(\Theta, x) = (\det \Theta)^{-1}\frac{\varphi(\Theta^{-1}x)f(\Theta^{-1}x)}{f(x)}$, et Θ une matrice carrée inversible de taille d.

Dans une procédure Monte Carlo, il est naturel de chercher un (le) paramètre θ qui minimise la variance de l'estimateur $g(\theta, X)$ ou de manière équivalente, le paramètre θ solution du problème

$$\min_{\theta} \mathbb{E}[g^2(\theta, X)]. \tag{.3.1}$$

Cela, dans le but de réduire l'erreur statistique de l'estimation. Si θ^* est un minimiseur de (.3.1) et si $(X_n)_{n\geq 0}$ est une suite de copies indépendantes de la loi de X alors on sait qu'il est souvent bénéfique d'utiliser l'approximation Monte Carlo classique (voir Chapitre 2)

$$\mathbb{E}[\varphi(X)] \simeq \frac{1}{N}\sum_{n=1}^{N} g(\theta^*, X_n).$$

Notre objectif dans ce chapitre est de développer un cadre rigoureux dans lequel, la quantité d'intérêt $\mathbb{E}[\varphi(X)]$ peut être approximée par

$$\mathbb{E}[\varphi(X)] \simeq \frac{1}{N}\sum_{n=1}^{N} g(\theta_n, X_n), \tag{.3.2}$$

pour N assez grand. $(\theta_n)_{n\geq 0}$ étant une suite (aléatoire) qui converge vers θ^*. L'originalité de cette méthode viendra du fait que nous obtenons pour cette estimation ergodique un Théorème Central Limite (TCL) pour lequel la variance théorique est estimable sur

l'échantillon. Nous verrons d'ailleurs que cette variance peut être estimée de manière empirique avec les mêmes trajectoires Monte Carlo $(X_n)_{n\geq 0}$ utilisées dans (.3.2). Ce qui constituera une méthode de Monte Carlo qui réduit sa propre variance de manière dynamique.

Remarque .27. Nous avons vu au Chapitre 1 que si le vecteur X suivait une loi de gauss standard, alors la transformation g donnée dans la remarque précédente devient

$$g(\theta, x) = \varphi(x+\theta)e^{-\theta\cdot x - \frac{1}{2}\|\theta\|^2}.$$

Dans cette écriture, $\|x\|$ représente la norme euclidienne d'un vecteur $x \in \mathbb{R}^d$ et $x \cdot y$ le produit scalaire euclidien de deux vecteurs $x, y \in \mathbb{R}^d$.

3.3 Principaux Résultats

Dans cette section nous énonçons et prouvons les résultats théoriques sur lesquels repose la mise en œuvre pratique de notre méthode. Nous nous plaçons donc dans le cadre général de l'estimation de

$$\mathbb{E}[\varphi(X)],$$

pour une fonction

$$\varphi : \mathbb{R}^d \to \mathbb{R}^p \quad \text{telle que} \quad \mathbb{E}[\|\varphi(X)\|] < +\infty.$$

La variable X suit une loi $d-$dimensionnelle définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose qu'il existe une transformation

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

qui vérifie les propriétés

$$(H_{11}) \quad \exists \quad 1 < a \leq 2, \forall\, \theta \in \mathbb{R}^q, \quad g(\theta, X) \in \mathbb{L}^{2a}(\mathbb{P}), \quad \text{et} \quad \mathbb{E}[g(\theta, X)] = \mathbb{E}[\varphi(X)].$$

Les résultats que nous avons obtenus sont les suivants.

Théorème .13. **Loi Forte des Grands Nombres.**
Pour un $\theta^* \in \mathbb{R}^q$, supposons qu'il existe une suite de variables aléatoires $(\theta_n)_{n\geq 0}$ definie sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que

$$\theta_n \underset{n\to+\infty}{\xrightarrow{p.s.}} \theta^* \qquad (\star).$$

Sous l'hypothèse (H_{11}), si nous supposons de plus que

(H_{12}) *La transformation* $\theta \longmapsto \mathbb{E}[\|g(\theta, X)\|^{2a}] := s_{2a}(\theta)$ *est continue en* θ^*
(H_{13}) $(\theta_n)_{n\geq 0}$ *est adaptée à la filtration engendrée par* θ_0 *et* $(X_n)_{n\geq 0}$, *et pour tout* $n \geq 0$, $\mathbb{E}(s_{2a}(\theta_n)) < +\infty$,

alors:

$$S_n := \frac{1}{n}\sum_{k=1}^{n} g(\theta_{k-1}, X_k) \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}[\varphi(X)],$$

avec $(X_n)_{n\geq 0}$ *une suite i.i.d. de copies de* X

Remarque .28. Il faut noter que dans le Théorème ci-dessus la suite $(\theta_n)_{n\geq 0}$ peut être déterministe ou stochastique.

Le théorème suivant nous précise la vitesse de convergence de l'estimation

Théorème .14. **Théorème Central Limite.**
Sous l'hypothèse (H_{11}) *et dans le cadre du Théorème .13, nous avons la convergence en loi suivante*

$$\sqrt{n}[S_n - \mathbb{E}[\varphi(X)]] \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^2)$$

avec

$$\Sigma^2 := \mathbb{E}\left[(g(\theta^*, X) - \mathbb{E}[\varphi(X)])\,(g(\theta^*, X) - \mathbb{E}[\varphi(X)])^T\right],$$

où u^T *désigne le transposé d'un vecteur* u *de* $\mathbb{R}^p$.

Ce qui induit le corollaire suivant

Corollaire .15. **Estimation de la variance empirique.**
Si l'hypothèse (H_{11}) *est vérifiée avec* $a = 2$ *et si*

$$\Sigma_n^2 := \frac{1}{n}\sum_{k=1}^{n} g(\theta_{k-1}, X_k) g(\theta_{k-1}, X_k)^T - S_n S_n^T$$

est une matrice inversible, alors

$$\Sigma_n^2 \xrightarrow[n\to+\infty]{p.s.} \Sigma^2 \quad et \quad \sqrt{n}\left(\Sigma_n^2\right)^{-1} \cdot (S_n - \mathbb{E}[\varphi(X)]) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, I_p).$$

Remarque .29. Il est important de noter que les variables aléatoires $(g(\theta_{n-1}, X_n))_{n\geq 1}$ ne sont ni indépendantes ni distribuées selon la même loi. Les Théorèmes .13 et .14 ci-dessus ne rentrent donc pas dans le cadre d'une méthode de Monte Carlo classique.

Remarque .30. Le Corollaire .15 montre que l'écart-type de l'estimation peut être estimé avec un coût supplémentaire négligeable en utilisant les mêmes trajectoires $(g(\theta_{n-1}, X_n))_{n\geq 1}$ déjà générées exactement comme dans une méthode de Monte Carlo standard. Cette faculté que possède la méthode de donner une estimation de l'erreur statistique avec un coût marginal, par rapport à l'estimation de la moyenne, est importante dans les applications.

Nous donnons maintenant une preuve de ces résultats. Par souci de simplicité, nous allons effectuer cette preuve pour $p = 1$, l'extension au cas $p > 1$ ne posant pas de problèmes.

Preuve. (du Théorème .13) La démonstration de ce théorème se fait par construction d'une martingale ayant de bonnes propriétés.
Considérons la suite $(M_n)_{n\geq 0}$ definie par $M_0 = 0$ et

$$M_n = \sum_{k=1}^{n} [g(\theta_{k-1}, X_k) - \mathbb{E}[\varphi(X)]], \quad n \geq 1.$$

Soit $\mathcal{F}_n = \sigma\{\theta_0, X_i, 0 \leq i \leq n\}$ la filtration propre de $(X_n)_{n\geq 0}$ (et de θ_0). Il est clair en utilisant l'hypothèse (H_{11}) que

$$\mathbb{E}[g(\theta_{n-1}, X_n)/\mathcal{F}_{n-1}] = \mathbb{E}[\varphi(X)],$$

et que $(M_n)_{n\geq 0}$ est une martingale par rapport à $(\mathcal{F}_n)_{n\geq 0}$. D'après (H_{13}), on a

$$\begin{aligned}
\mathbb{E}((\Delta M_n)^2) &= \mathbb{E}[\mathbb{E}[(g(\theta_{n-1}, X_n) - \mathbb{E}[\varphi(X)])^2/\mathcal{F}_{n-1}]] \\
&= \mathbb{E}[\mathbb{E}(g^2(\theta_{n-1}, X_n)/\mathcal{F}_{n-1})] - \mathbb{E}[\varphi(X)]^2 \\
&\leq \mathbb{E}[s_2(\theta_{n-1})] < +\infty,
\end{aligned}$$

et $(M_n)_{n\geq 0}$ est une $\mathcal{F}_n-$martingale de carré intégrable. Son processus crochet oblique est donné par

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

Comme $\theta_n \xrightarrow{p.s.} \theta^*$, le Lemme .9 de Césaro implique

$$\frac{\langle M \rangle_n}{n} \underset{n}{\longrightarrow} s_2(\theta^*) - \mathbb{E}[\varphi(X)]^2 = \mathbb{V}\mathrm{ar}(g^2(\theta^*, X)) = \sigma^2 \quad p.s.$$

Le Théorème .4 s'applique et donne

$$\frac{M_n}{n} \underset{n}{\longrightarrow} 0 \quad p.s.$$

C'est-à-dire, le résultat recherché. □

Preuve. (du Théorème .14). En tenant compte du résultat du Théorème .13, il suffit de démontrer que la condition de Lindberg est vérifiée pour la martingale $(M_n)_{n\geq 0}$. Or, un calcul simple mène à l'estimation suivante

$$\begin{aligned}\mathbb{E}\Big(|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))|^{2a} / \mathcal{F}_{k-1} \Big) &\leq 2^{2a-1} \Big(\mathbb{E}[|g(\theta_{k-1}, X_k)|^{2a} / \mathcal{F}_{k-1}] \\ &\qquad + |\mathbb{E}(\varphi(X))|^{2a} \Big) \\ &= 2^{2a-1}(s_{2a}(\theta_{k-1}) + |\mathbb{E}(\varphi(X))|^{2a}).\end{aligned}$$

L'hypothèse (H_{12}) et le Lemme de Césaro montrent alors que

$$\frac{1}{n}\sum_{k=1}^{n}(s_{2a}(\theta_{k-1}) + |\mathbb{E}(\varphi(X))|^{2a}) \underset{n}{\longrightarrow} L,$$

avec

$$L = s_{2a}(\theta^*) + |\mathbb{E}(\varphi(X))|^{2a},$$

une constante réelle positive pour laquelle nous avons

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\Big(|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))|^{2a} / \mathcal{F}_{k-1} \Big) \leq L \quad p.s.$$

Maintenant pour un $b > 0$ fixé et pour tout $n \geq 0$ si on définit

$$F_n(b) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\Big(|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))|^2 1_{\{|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))| > b\}} / \mathcal{F}_{k-1} \Big),$$

alors on a certainement l'estimation

$$F_n(b) \leq \frac{b^{-2(a-1)}}{n}\sum_{k=1}^{n} \mathbb{E}\Big(|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))|^{2a} / \mathcal{F}_{k-1} \Big) \quad p.s.$$

En faisant varier b à travers la suite $b_n = \varepsilon\sqrt{n}$ pour un $\varepsilon > 0$ arbitraire on obtient

$$F_n(b_n) \leq \frac{\varepsilon^{-2(a-1)}}{n^a} \sum_{k=1}^{n} \mathbb{E}\Big(|g(\theta_{k-1}, X_k) - \mathbb{E}(\varphi(X))|^{2a} / \mathcal{F}_{k-1} \Big) \quad p.s.$$

Ce qui prouve que

$$\limsup_{n\to+\infty} F_n(b_n) = \lim_{n\to+\infty} F_n(b_n) = 0, \quad p.s.$$

Ainsi la condition de Lindberg est vérifiée pour la martingale $(M_n)_{n\geq 0}$. Finalement le résultat voulu s'obtient en appliquant le Théorème .12. □

Preuve. (du Corollaire .15). Pour prouver le corollaire, nous avons simplement besoin de montrer que les deux suites suivantes

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

convergent presque sûrement vers la même limite. Pour cela on définit par récurrence la suite

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

Nous rappelons que dans le Corollaire .15, nous avons renforcé l'hypothèse (H_{11}) en supposant que $a = 2$. Dans ce cas une argumentation similaire à celle de la preuve du Théorème .13 montre que $(\overline{M}_n)_{n\geq 0}$ est une martingale de carré intégrable par rapport à la filtration propre de $(X_n)_{n\geq 0}$ (et de θ_0) en l'occurence $(\mathcal{F}_n)_{n\geq 0}$. Son processus crochet est donné par

$$\begin{aligned}\langle \overline{M} \rangle_n &= \sum_{k=1}^{n} \Big(\mathbb{E}[g^4(\theta_{k-1}, X_k)/\mathcal{F}_{k-1}] - s_2^2(\theta_{k-1}) \Big) \\ &= \sum_{k=1}^{n} \Big(s_4(\theta_{k-1}) - s_2^2(\theta_{k-1}) \Big).\end{aligned}$$

De nouveau d'après le Lemme de Césaro et la continuité des fonctions s_4 et s_2 on voit que

$$\frac{\langle \overline{M} \rangle_n}{n} \underset{n}{\longrightarrow} s_4(\theta^*) - s_2^2(\theta^*) = \mathbb{V}\mathrm{ar}(g^2(\theta^*, X)) > 0 \quad p.s.$$

La première partie du Théorème .4 s'applique et montre que

$$\frac{\overline{M}_n}{n} \underset{n}{\longrightarrow} 0,$$

ce qu'on réecrit sous la forme

$$\left[\frac{1}{n}\sum_{k=1}^{n} g^2(\theta_{k-1}, X_k) - \frac{1}{n}\sum_{k=1}^{n} s_2(\theta_{k-1})\right] \underset{n}{\longrightarrow} 0.$$

Mais comme $\frac{1}{n}\sum_{k=1}^{n} s_2(\theta_{k-1}) \underset{n}{\longrightarrow} s_2(\theta^*)$ on obtient finalement

$$\sigma_n^2 = \frac{1}{n}\sum_{k=1}^{n} g^2(\theta_{k-1}, X_k) - S_n^2 \underset{n}{\longrightarrow} s_2(\theta^*) - s_1^2(\theta^*) = \mathbb{V}\mathrm{ar}(g(\theta^*, X)).$$

Ce qui complète notre preuve. □

Remarque .31. Si la suite $(\theta_n)_{n\geq 0}$ est ergodique, on a toujours $\frac{\langle M\rangle_n}{n} \underset{n}{\longrightarrow} s_2(\theta^*) - \mathbb{E}[\varphi(X)]^2$, et la Loi Forte des Grands Nombres reste vraie. Pour le Théorème Central Limite, c'est un peu plus compliqué. Mais cette remarque justifie l'utilisation pratique des algorithmes stochastiques θ_n à pas constant pour faire de la réduction de variance.

Dans la section qui suit, notre objectif est de montrer dans quelle mesure ces résultats théoriques peuvent permettre de mettre en œuvre une technique de réduction de variance Monte Carlo.

3.4 Une procédure de minimisation de variance

Supposons comme au debut de la section précedente que la quantité $\mathbb{E}[\varphi(X)]$ peut se mettre sous la forme $\mathbb{E}[g(\theta, X)]$. Comme nous l'avons vu au Chapitre 2, il est naturel de déterminer l'estimateur $g(\theta^*, X)$ dont la variance est la plus petite. Ce qui revient à déterminer le paramètre θ^* qui vérifie

$$\min_{\theta} V(\theta) = V(\theta^*),$$

avec

$$V(\theta) = \mathbb{E}[g^2(\theta, X)].$$

Sous certaines hypothèses, $\nabla V(\theta)$ existe et peut être écrit sous forme d'une espérance

$$\nabla V(\theta) = \mathbb{E}[F(\theta, X)], \quad \theta \in \mathbb{R}^d,$$

qui vérifie

$$\mathbb{E}[F(\theta^*, X)] = 0.$$

Lorsque cette espérance n'est pas accessible numériquement, mais qu'on sait simuler la loi de X, une idée naturelle est d'utiliser une version adéquate des algorithmes de Robbins-Monro pour résoudre le problème ci-dessus.
Supposons pour l'instant qu'il existe une telle version des algorithmes de Robbins-Monro que nous notons $(\theta_n)_{n\geq 0}$ et qui vérifie

$$\theta_n \xrightarrow[n\to+\infty]{p.s.} \theta^*.$$

Supposons que la suite approximante $(\theta_n)_{n\geq 0}$ est générée à partir d'une suite $(X_n)_{n\geq 0}$ de copies indépendantes de la loi de X. Alors la méthode d'estimation de l'espérance $\mathbb{E}[\varphi(X)]$, que nous proposons ici, se résume comme suit:

1. Le Théorème .13 affirme que

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} g(\theta_{n-1}, X_n) = \mathbb{E}[\varphi(X)]$$

   où $(X_n)_{n\geq 0}$ est la même suite utilisée pour générer $(\theta_n)_{n\geq 0}$.

2. D'après le Théorème .14, la variance théorique du théorème de la limite centrale pour ce problème est donnée par :

$$\sigma^2 = \mathbb{V}\mathrm{ar}[g(\theta^*, X)] = \min_{\theta} \mathbb{V}\mathrm{ar}[g(\theta, X)].$$

   Cette variance est optimale au sens où, c'est la plus petite variance que l'on peut espérer avoir dans cette formulation paramétrique du problème.

3. Finalement en utilisant le Corollaire .15 on a l'estimation empirique suivante de la variance théorique

$$\sigma_N^2 = \frac{1}{N} \sum_{n=1}^{N} g^2(\theta_{n-1}, X_n) - S_N^2.$$

   Cette estimation empirique est obtenue en utlisant les mêmes trajectoires $(X_n)_{n\geq 0}$ générées précédemment, et c'est elle qui permet d'avoir des estimations d'erreur à l'aide d'intervalle de confiance.

Remarque .32. La procédure décrite ci-dessus est similaire en tout point à la procédure Monte Carlo standard sauf pour la prise en compte additionnelle de la simulation de la suite approximante $(\theta_n)_{n\geq 0}$. Mais ce calcul supplémentaire qui se fait avec les mêmes trajectoires déjà simulées permet en retour, une réduction de variance au fur et à mesure des itérations Monte Carlo de l'estimation.

Nous détaillons dans la suite deux applications possibles de la méthode. La première (et la plus importante) concerne la finance numérique, et la deuxième traite d'applications en fiabilité des systèmes.

3.4.1 Applications financières

Nous rappelons que le prix d'arbitrage d'une option qui paie le montant $\psi(S_t, t \leq T)$ à la date T est donné par (voir par exemple [HUL03])

$$c = \mathbb{E}[e^{-\int_0^T r(t)dt}\psi(S_t, t \leq T)], \tag{.3.3}$$

où le sous-jacent S est supposé suivre la dynamique suivante

$$dS_t = S_t(r(t)dt + \sigma(t, S_t)dW_t),\ S_0 = x, \tag{.3.4}$$

avec $r(t)$ le taux d'intérêt sans risque, $\sigma(t, y)$ la fontion de volatilité locale, W un mouvement brownien et $x > 0$ fixé.
Lorsqu'on ne dispose pas d'une solution explicite de la diffusion (.3.4), comme au chapitre 2, nous considérons disposer d'une discrétisation acceptable de cette solution sur une grille de temps $0 = t_0 < t_1 < \cdots < t_d = T$. Dans ce cas en pratique, pour estimer la valeur de c, on est réduit à calculer

$$\hat{c} = \mathbb{E}[e^{-\sum_{i=1}^d r(t_i)}\hat{\psi}(S_{t_1}, \ldots, S_{t_d})].$$

$\hat{c}$ peut être réecrit sous la forme

$$\hat{c} = \mathbb{E}[\varphi(X)],$$

où $X = (X_1, \ldots, X_d) \sim \mathcal{N}(0, I_d)$ et I_d est la matrice identité de $\mathbb{R}^d$. φ est une transformation dont l'expression dépend à la fois de la discrétisation de S et du payoff ψ. Nous avons vu à la section 3 du chapitre 2 que si g désigne la fonction définie par

$$g(\theta, x) = \varphi(x + \theta)e^{-\theta \cdot x - \frac{1}{2}\|\theta\|^2},$$

alors

$$\hat{c} = \mathbb{E}[g(\theta, X)]$$

et la variance est donnée par

$$V(\theta) = \mathbb{E}[g^2(\theta, X)] - \hat{c}^2.$$

Nous avons également vu dans cette même section que le problème de minimisation

$$\min_{\theta \in \mathbb{R}^d} V(\theta),$$

possède une unique solution notée θ^* et que θ^* était aussi la seule solution du problème

$$\nabla V(\theta) := \mathbb{E}[(\theta - X)\varphi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}] = 0. \tag{.3.5}$$

Nous avons enfin démontrer dans ce même chapitre (voir le Théorème .12) que θ^* est limite presque sûre de l'algorithme stochastique (.1.11-.1.13) et la condition $(\star)$ du Théorème .13 est satisfaite.
Nous possédons donc une suite approximante et sommes en mesure de développer notre procédure de réduction de variance. La proposition suivante résume tout cela.

Proposition 3.4.1. *Si φ vérifie $\mathbb{E}(|\varphi(X)|^{4\delta}) < +\infty$ pour un certain $\delta > 1$, et si nous notons $g(\theta, x) = \varphi(x+\theta)e^{-\theta\cdot x-\frac{1}{2}\|\theta\|^2}$, alors les hypothèses (H_{11}), (H_{12}) et (H_{13}) sont satisfaites et les Théorèmes .13 et .13 ainsi que le Corollaire .15 s'appliquent.*

Preuve. C'est facile de vérifier que (H_{11}) est satisfaite avec $a = 2$. De plus on a

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

si s_p désigne le moment d'ordre p de $g(\theta, X)$. Maintenant supposons que $\|\theta\| \leq C$ avec C une constante réelle positive strictement. Avec la notation

$$v_p(\theta, x) \ = |\varphi(x)|^p e^{-(p-1)\theta\cdot x+\frac{p-1}{2}\|\theta\|^2},$$

on a l'estimation

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

En utilisant l'inégalité de Hölder il vient

$$\begin{aligned}\int |\varphi(x)|^p e^{C(p-1)\|x\|}e^{-\frac{1}{2}\|x\|^2}dx \leq \left(\int e^{\frac{\delta C}{\delta-1}(p-1)\|x\|}e^{-\frac{1}{2}\|x\|^2}dx\right)^{1-\frac{1}{\delta}}\left(\int |\varphi(x)|^{\delta p}\times\right.\\ \left.\times\, e^{-\frac{1}{2}\|x\|^2}dx\right)^{\frac{1}{\delta}}.\end{aligned} \tag{.3.6}$$

Mais $\mathbb{E}(|\varphi(X)|^{4\delta}(X)) < \infty$, avec $\delta > 1$, et le Théorème de Lebesgue entraine que la fonction s_p est continue pour $p = 2, 3, 4$. L'hypothèse (H_{12}) est donc satisfaite.
Par définition même de l'algorithme (.1.11-.1.13), on a pour $n \geq 0$, $\theta_n \leq U_n$, $(U_n)_{n\geq 0}$ faisant référence à la suite de projection intervenant dans la définition de cet algorithme. Et d'après (.3.6) on a pour chaque $n \geq 0$ fixé

$$s_p(\theta_n) \leq C_n, \quad \text{pour} \quad p = 2, 3, 4,$$

et une certaine constante strictement positive C_n. L'hypothèse (H_{13}) s'applique et on obtient le résultat énoncé. □

La sous-section suivante traite des possibilités d'applications en fiabilité.

3.4.2 Un exemple en fiabilité

En fiabilité des systèmes, l'on a souvent besoin de modéliser la durée de vie T (aléatoire) d'un composant puis d'estimer sa durée de vie moyenne. La variable T est suposée posséder une fonction de distribution continue $F(t) = \mathbb{P}(T \leq t),\ t \geq 0$. La plupart du temps les données empiriques sur le fonctionnement des composants d'un système sont recueillies en termes de proportion conduisant à l'introduction d'un taux de défaillance noté $\lambda(t)$. Ce taux mesure à l'instant t la proportion de composants qui tombent en panne avant l'instant $t + dt$ sachant qu'ils étaient en fonction après la date t. Sa définition est donc donnée par

$$\lambda(t) = \lim_{h \searrow 0} \frac{1}{h} \mathbb{P}(T \leq t + h / T > t).$$

On définit aussi le taux de défaillance cumulée $\Lambda(t)$ (ou fonction de hasard) par $\Lambda(t) = \int_0^t \lambda(s)ds$. Il est bien connu que (voir par exemple [AJ99]) la fonction de répartition F est déterminée de manière unique par la relation

$$F(t) = 1 - e^{-\Lambda(t)}, \quad t \geq 0.$$

Supposons que l'on soit intéressé par le calcul de la durée de vie moyenne du système $\mathbf{MTTF} = \mathbb{E}(T)$. La méthode la plus simple pour calculer **MTTF** est de faire l'hypothèse que l'on connait la loi de T. En pratique une hypothèse de loi exponentielle peut s'avérer satisfaisante (c'est le cas, par exemple, des composants électroniques neufs). Ici nous décidons de ne faire aucune hypothèse à priori sur la loi de T.
Soit ψ l'inverse (ou pseudo-inverse généralisée) de Λ. Si X est une variable de loi exponentielle de moyenne 1, alors les variables aléatoires $\psi(X)$ et T sont distribuées selon la même loi. En effet

$$\mathbb{P}(\psi(X) > t) = \mathbb{P}(X > \Lambda(t)) = e^{-\Lambda(t)} = \mathbb{P}(T > t).$$

On renforce l'hypothèse (H_{11}) de la manière suivante

$$(H_{14}): \quad \mathbb{P}(\psi(X) > 0) > 0,\ \exists\, \varepsilon^* > 0, \text{ et } a > 1 \text{ tels que } \int_{\mathbb{R}_+} x^a \psi^{2a}(x) e^{-\varepsilon^* x} dx < +\infty.$$

Remarque .33. Cette condition peut paraître restrictive, mais il n'en est rien car on peut montrer que $\lim_{x \to +\infty} \psi(x) = +\infty$. En général, la fonction ψ croit vers l'infini comme une fonction puissance $x \mapsto x^\alpha$, $\alpha > 0$, et l'hythotèse (H_{14}) est vérifiée pour ε^* arbitrairement petit. C'est par exemple le cas, lorsque la variable T suit une loi de Weibull, de Fréchet ou encore une loi de Gumbel.

Notre objectif est de montrer que l'on peut mettre en œuvre une méthode de Monte Carlo efficace pour l'estimation de **MTTF** lorsqu'il n'est pas facilement calculable mais qu'une "bonne approximation numérique" de la fonction ψ est disponible.
Un simple changement de variables donne

$$\mathbf{MTTF} = \mathbb{E}[\psi(X)] = \mathbb{E}[\psi(X/\theta)e^{(1-1/\theta)X}]/\theta, \quad \theta > 0.$$

Comme dans la section précedente, il est naturel de déterminer si la variance de la famille d'estimateurs $(g(\theta, X))_{\theta>0}$ possède un minimum atteint en un point θ^*. Pour cela, remarquons d'abord que

$$\mathbb{V}\mathrm{ar}\left(\psi(X/\theta)e^{(1-1/\theta)X}/\theta\right) < +\infty \quad \Longleftrightarrow 0 < \theta < 2.$$

Nous devons donc nous restreindre au problème de minimisation suivant

$$\min_{0<\theta<2} h(\theta), \quad \text{avec} \quad h(\theta) = \mathbb{E}[\psi^2(X/\theta)e^{2(1-1/\theta)X}]/\theta^2. \tag{.3.7}$$

La peuve de la proposition suivante est analogue à celle de 2.3.1.

Proposition 3.4.2. *Sous l'hypothèse* (H_{14})*, la fonction* h *définie ci-dessus est strictement convexe sur* $]0,\ 2[$ *et vérifie*

$$\lim_{\theta\to\{0,2\}} h(\theta) = +\infty.$$

h *atteint donc son minimum en un unique point* $\theta^* \in]0,\ 2[$*. De plus* θ^* *est donnée comme unique solution du problème*

$$H(\theta) := \mathbb{E}[(\theta X - 1)\psi^2(X)e^{(\theta-1)X}] = 0, \quad 0 < \theta < 2.$$

Remarque .34. Une transformation simple permet de voir que

$$h(\theta) = \mathbb{E}[\psi^2(X)e^{(\theta-1)X}]/\theta, \quad 0 < \theta < 2.$$

Pour des raisons techniques, nous devons nous limiter au problème suivant

$$\min_{\gamma^*\le\theta\le 2-\delta^*} h(\theta), \quad \text{avec} \quad \gamma^* > 0 \quad \text{et} \quad a\delta^* = a - 1 + \varepsilon^*. \tag{.3.8}$$

Cette restriction est une sorte de localisation du problème de départ. Elle nous permet de traiter jusqu'au bout notre problème de minimisation par l'utilisation d'un résultat dû à Bernard Delyon (cf. [DEL00]).

Remarque .35. Le choix de δ^* apparaît clairement dans une preuve que nous donnons ci-dessous. celui de γ^* est moins naturel, mais il est essentiel pour des "besoins de démonstration". Comme le cas $\theta = 1$ correspond à la méthode de Monte Carlo standard, à partir du moment où on choisit δ^* et γ^* tels que $1 \in [\gamma^*,\ 2 - \delta^*]$, la procédure que nous présentons devrait toujours permettre de réduire la variance de l'estimation.

D'après la Proposition 3.4.2 et la remarque ci-dessus, le problème de minimisation à résoudre est le suivant

$$H(\theta) = 0, \quad \gamma^* \leq \theta \leq 2 - \delta^*. \tag{.3.9}$$

Son unique solution notée $\hat{\theta}$ coincidera avec θ^* si δ^* et γ^* sont suffisamment petits. Dans ce qui suit, nous nous concentrons sur la recherche d'une suite approximante du paramètre $\hat{\theta}$.
On considère la suite $(\theta_n)_{n\geq 0}$ définie par la donnée de $\theta_0 \in [\gamma^*, 2-\delta^*]$ et par la relation de récurrence pour $n \geq 0$

$$\theta_{n+1} = \begin{cases} \theta_n - \gamma_{n+1} Y_{n+1} & \text{si} \quad \theta_n - \gamma_{n+1} Y_{n+1} \in [\gamma^*, 2-\delta^*], \\ \theta_n & \text{sinon,} \end{cases} \tag{.3.10}$$

avec $(\gamma_n)_{n\geq 0}$ une suite réelle positive vérifiant

$$\sum_{n\geq 0} \gamma_n = +\infty \quad \text{et} \quad \sum_{n\geq 0} \gamma_n^a < +\infty. \tag{.3.11}$$

La suite Y_n est donnée d'après l'expression de $H(\theta)$ par

$$Y_{n+1} = (\theta_n X_{n+1} - 1)\psi^2(X_{n+1}) e^{(\theta_n - 1)X_{n+1}}, \tag{.3.12}$$

où $(X_n)_{n\geq 0}$ est une suite de copies indépendantes de la loi exponentielle de moyenne 1. La proposition suivante assure que la version des algorithmes stochastiques définie ci-dessus constitue bien une suite approximante (aléatoire) du paramètre $\hat{\theta}$ recherché.

Proposition 3.4.3. *Sous l'hypothèse* (H_{14}) *si on note*

$$e_{n+1} = H(\theta_n) - Y_{n+1} \text{ et } r_{n+1} = \gamma_{n+1} Y_{n+1} \mathbf{1}_{\{\theta_n - \gamma_{n+1} Y_{n+1} \notin [\gamma^*, 2-\delta^*]\}}$$

alors nous avons les résultats suivants:

$$(R_1) : H \text{ est continue sur }]0,\, 2[\text{ et } (x-\theta^*)H(x) > 0, \ \forall x \in]0,\, 2[- \{\theta^*\}.$$

$$(R_2) : \lim_{p\to+\infty} \sum_{n=1}^{p} \gamma_n e_n < +\infty \ p.s. \quad \text{et} \quad \lim_{n\to+\infty} r_n = 0 \ p.s.$$

L'algorithme (.3.10-.3.12) converge donc presque sûrement vers l'unique solution θ^* *du problème .3.9.*

Preuve. L'assertion (R_1) est une conséquence immédiate de la convexité stricte de la fonction h (voir la Proposition 3.4.2). Soit $(\bar{\mathcal{F}}_n)_{n\geq 0}$ la filtration propre de θ_0 et de

$(X_k)_{k\geq 0}$. Il est alors clair que $\mathbb{E}[Y_{n+1}/\bar{\mathcal{F}}_n] = H(\theta_n)$. De plus on a

$$\begin{aligned}
\mathbb{E}[|e_{n+1}|^a/\bar{\mathcal{F}}_n] &\leq \mathbb{E}[(|H(\theta_n)| + |Y_{n+1}|)^a/\bar{\mathcal{F}}_n] \\
&\leq 2^{a-1}\Big(|H(\theta_n)|^a + \mathbb{E}[|Y_{n+1}|^a/\bar{\mathcal{F}}_n]\Big) \\
&\leq 2^a s_a(\theta_n), \quad \text{avec} \quad s_a(\theta) = \mathbb{E}[|\theta X - 1|^a \psi^{2a}(X) e^{a(\theta-1)X}] \\
&\leq \mathbb{E}[((2-\delta^*)X+1)^a \psi^{2a}(X) e^{a(1-\delta^*)X}] \quad \text{car} \quad \theta_n \leq 2-\delta^* \\
&= \mathbb{E}[((2-\delta^*)X+1)^a \psi^{2a}(X) e^{(1-\varepsilon^*)X}] \\
&< +\infty \; p.s. \quad \text{grace à} \quad (H_{14}).
\end{aligned}$$

Par conséquent la suite $(M_n)_{n\geq 0}$ définie par $M_n = \sum_{k=1}^n \gamma_k e_k$, et $M_0 = 0$ est une $\bar{\mathcal{F}}_n$-martingale de puissance $a-$intégrable, telle que

$$\sum_{k=1}^n \mathbb{E}[|M_k - M_{k-1}|^a/\bar{\mathcal{F}}_{k-1}] = \sum_{k=1}^n \gamma_k^a \mathbb{E}[|e_k|^a/\bar{\mathcal{F}}_{k-1}] < +\infty \quad p.s.$$

D'après le Théorème de Chow .3 on a

$$\sum_{n=1}^{+\infty} \gamma_n e_n < +\infty \quad p.s.$$

Or par définition même la suite $(H(\theta_n))_{n\geq 0}$ est bornée; donc

$$\lim_{n\to+\infty} \gamma_n Y_n = 0 \; p.s., \quad \text{et} \quad \lim_{n\to+\infty} r_n = 0 \; p.s.$$

Pour conclure, il suffit de réecrire l'algorithme (.3.6) sous la forme

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n) + \gamma_{n+1} e_{n+1} + \gamma_{n+1} r_{n+1},$$

et d'appliquer le résultat du Théorème .7. □

Nous terminons cette section avec la proposition suivante dont la preuve est analogue à celle de la proposition 3.4.1.

Proposition 3.4.4. *Si l'hypothèse* (H_{14}) *est satisfaite avec le paramètre* $a = 2$, *alors les conditions* (H_{11}), (H_{12}) *and* (H_{13}) *sont satisfaites avec* $g(\theta, x) = \frac{1}{\theta}\psi(\frac{x}{\theta})e^{(1-\frac{1}{\theta})x}$ *et les Théorèmes .13, .14 et Corollaire .15 s'appliquent.*

3.5 Illustrations numériques

Nous allons présenter deux applications numériques:

1. des cas gaussiens venant de la finance, la méthode que nous proposons consiste à approximer la quantité $\mathbb{E}[\varphi(X)]$ par

$$\frac{1}{n}\sum_{k=1}^{n}\varphi(X_k+\theta_{k-1})e^{-\theta_{k-1}\cdot X_k-\frac{1}{2}\|\theta_{k-1}\|^2}, \qquad ((\star)_a)$$

   avec $(\theta_k)_{k\geq 0}$ défini par l'algorithme (.1.11-.1.13).

2. Dans l'exemple de fiabilité de la section précédente, on approxime $\mathbb{E}[\psi(X)]$ (X suit une loi exponentielle de moyenne 1) par

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\theta_{k-1}}\psi\left(\frac{X_k}{\theta_{k-1}}\right)e^{X_k-\frac{X_k}{\theta_{k-1}}}, \qquad ((\star)_b)$$

   où $(\theta_k)_{k\geq 0}$ est donné par (.3.10-.3.12).

D'après les résultats théoriques vus plus haut, ces estimations permettent de réduire la variance statistique par rapport à une estimation Monte Carlo standard. La complexité de notre méthode est de l'ordre de $C\times n\times d$ alors que celle d'une méthode de Monte Carlo classique est de $C^{'}\times n\times d$, d étant la taille du vecteur X et C et $C^{'}$ deux constantes réelles positives. L'appel itératif à la fonction *exponentielle* dans l'implémentation des approximations ci-dessus suggère que C va être légèrement supérieur à $C^{'}$. Néanmoins nous verrons que le gain *in fine* de la méthode justifie largement cette légère augmentation de sa complexité de calcul.

3.5.1 Le modèle de Black et Scholes

Notre premier exemple concerne le pricing Monte Carlo d'un Call européen dans le modèle de Black et Scholes. Ce modèle est defini par (.3.4) dans le cas particulier d'un taux d'intérêt r et d'une volatilité $\sigma>0$ constants. La simulation exacte du sous-jacent se fait alors par la formule

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t}.$$

Le prix d'arbitrage d'un Call (*resp.* Put) est donné par

$$e^{-rT}\mathbb{E}[\max(S_T-K,0)] \quad (resp.\ e^{-rT}\mathbb{E}[\max(K-S_T,0)]).$$

C'est un fait bien connu que la variance Monte Carlo d'un Put est souvent inférieure à celle d'un Call. Nous rappelons la formule dite de parité Call–Put

$$e^{-rT}\mathbb{E}[\max(S_T - K, 0)] - e^{-rT}\mathbb{E}[\max(K - S_T, 0)] = S_0 - Ke^{-rT}.$$

D'après cette formule et la remarque faite précedemment, il est souvent plus avantageux, lorsque l'on veut "pricer" un Call par une méthode de Monte Carlo, de pricer systématiquement le Put correspondant. Le but, ici, est de combiner notre méthode de Monte Carlo adaptative avec cette technique de parité Call–Put afin d'obtenir une réduction de variance encore plus importante. Cette idée est testée sur un Call européen de maturité $T = 1an$, sur un sous-jacent de valeur spot $S_0 = 100$, de volatilité $\sigma = 0.2$ ou 0.3 avec un taux d'intérêt sans risque $r = 0.05$ ou 0.1. Les résultats sont reproduits dans le Tableau 1. Err désigne l'erreur statistique sur les prix estimés et est défini par $\mathsf{Err} = \sigma_N/\sqrt{N}$, où σ_N est l'écart type Monte Carlo de l'estimation. Dans ce tableau, BS représente le prix Black & Scholes et RV désigne le rapport des variances. On y voit clairement que la variance Monte Carlo peut être réduite de façon très significative.
Cette observation plaide pour un usage (chaque fois que cela est possible) de la méthode adaptative même si une autre méthode de réduction de variance est déjà utilisée.

Tableau 1 Réduction de Variance estimée pour Call Européen loin dans la monnaie sous le modèle de Black–Scholes

		IS + Indirect				**Indirect**			**BS**	
	α	**K/x**	**Prix**	**Err**	**RV**	**Prix**	**Err**	**RV**	**Prix**
	5×10^4	0.7	33.540	0.0006	23751	33.542	0.005	382	33.540
$r = 0.05$	1×10^4	0.8	24.587	0.003	1131	24.585	0.012	51	24.589
$\sigma = 0.2$	1×10^3	0.9	16.698	0.007	118	16.689	0.023	11	16.699
	8×10^2	1.0	10.449	0.014	22	10.433	0.039	3	10.451
	5×10^4	0.7	36.722	0.0003	100793	36.724	0.003	877	36.722
$r = 0.1$	1×10^4	0.8	27.992	0.002	3371	27.994	0.009	105	27.993
$\sigma = 0.2$	1×10^3	0.9	19.986	0.005	293	19.980	0.018	20	19.989
	8×10^2	1.0	13.263	0.011	47	13.265	0.031	5.4	13.270
	5×10^4	0.7	37.320	0.003	2726	37.319	0.013	112	37.321
$r = 0.1$	1×10^4	0.8	29.428	0.006	424	29.420	0.023	31	29.432
$\sigma = 0.3$	1×10^3	0.9	22.503	0.012	101	22.500	0.036	11	22.510
	8×10^2	1.0	16.730	0.02	29	16.733	0.05	4.6	16.734

Les résultats sont basés sur 50,000 tirages Monte Carlo et Les autres paramètres de l'option sont $\mathbf{x = 100}$, $\mathbf{T = 1.0}$.

3.5.2 Le modèle de Heston

Nous appliquons maintenant la procédure au modèle à volatilité stochastique de Heston (voir Chapitre 2) dont la discrétisation s'écrit

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

avec $(X^i)_{i\geq 1}$ une suite de copies indépendantes de la loi normale centrée réduite. Comme nous l'avons vu au chapitre précedent, sous ce modèle, les prix des Call et Put européens sont disponibles sous forme analytique ([HES93]). Nous avons de nouveau implémenté ces formules. Les résultats obtenus sont dans le Tableau 2. **Q-CF** désigne le prix calculé avec cette formule, et **Brute** représente le prix et l'erreur statistique de la méthode de Monte Carlo standard. Une fois de plus la réduction de variance obtenue par notre méthode seule est significative. Mais pour souligner encore une fois l'intérêt de combiner notre méthode avec une autre technique de réduction de variance (plus ou moins standard), nous utilisons de nouveau la relation de parité Call–Put[1].

Tableau 2 Réduction de Variance estimée pour un Call Européen loin en dehors de la monnaie sous le modèle à volatilité stochastique de Heston

	IS					**Brute**		**Q-CF**
$\mathbf{v_0}$	$\boldsymbol{\alpha}$	$\mathbf{K/S_0}$	**Price**	**Err**	**VR**	**Price**	**Err**	**Price**
	200	1.4	0.37	0.006	18.2	0.40	0.03	0.38
	80	1.3	0.70	0.01	13.1	0.73	0.04	0.70
0.04	50	1.2	1.38	0.01	10.1	1.43	0.05	1.39
	40	1.1	4.93	0.02	8.0	3.03	0.06	2.94
	30	1.05	4.42	0.03	6.2	4.57	0.07	4.44
	80	1.4	1.00	0.01	13.5	1.04	0.05	1.01
	80	1.3	1.66	0.02	12.1	1.72	0.06	1.68
0.09	50	1.2	2.85	0.02	9.4	2.95	0.07	2.87
	40	1.1	4.97	0.03	8.4	5.14	0.09	5.04
	30	1.05	6.61	0.04	8.0	6.84	0.1	6.72

Nous avons utilisé 20,000 Tirages Monte Carlo et $\mathbf{n = 200}$ pas de discretisation. Les paramètres du Call sont $\mathbf{S_0 = 100}$, $\mathbf{r = 0.1}$, $\mathbf{T = 0.5}$, $\mathbf{k = 2}$, $\mathbf{\bar{v} = 0.01}$, $\sigma = \mathbf{0.5}$.

[1]Nous rappelons que la relation de parité Call–Put est une conséquence de l'absence d'opportunité d'arbitrage et de ce fait ne dépend pas du modèle choisi pour le sous-jacent

Le Tableau 3 contient les résultats que nous avons obtenus pour un Call européen loin dans la monnaie dans le modèle de Heston. On y voit que la variance de l'estimation peut être énormément réduite.
Les résultats numériques regroupés dans le Tableau 4 sont relatifs à une option d'achat sur moyenne arithmérique (Call Asiatique discret) dont le payoff s'écrit

$$\psi(S_{t_i},\, i=1...d) = \left(\frac{1}{d}\sum_{i=1}^{d} S_{t_i} - K\right)_+ .$$

Tableau 3 Réduction de Variance estimée pour un Call Européen loin dans la monnaie sous le modèle à volatilité stochastique de Heston

			IS + Indirect				**Indirect**		**Q-CF**
$\mathbf{v_0}$	$\boldsymbol{\alpha}$	$\mathbf{K/S_0}$	**Price**	**Err**	**VR**	**Price**	**Err**	**VR**	**Price**
	400	0.7	33.416	0.0007	18989	33.416	0.001	6609	33.417
	200	0.8	23.925	0.002	1437	23.927	0.003	793	23.933
0.04	200	0.9	14.641	0.007	173	14.653	0.01	75	14.650
	100	1.0	6.875	0.017	23	6.913	0.028	8.8	6.868
	25	1.05	4.449	0.027	6.6	4.484	0.039	3	4.440
	800	0.7	33.432	0.002	5833	33.435	0.003	2068	33.443
	400	0.8	24.070	0.005	632	24.088	0.009	212	24.086
0.09	100	0.9	15.476	0.015	70	15.498	0.023	29	15.485
	10	1.0	8.953	0.034	11	8.971	0.045	6	8.966
	10	1.05	6.701	0.047	4.6	6.714	0.057	3.1	6.718

Nous avons utilisé 20,000 Tirages Monte Carlo et $\mathbf{n = 200}$ pas de discretisation. Les paramètres du Call sont$\mathbf{S_0 = 100},\ \mathbf{r = 0.1},\ \mathbf{T = 0.5},\ \mathbf{k = 2},\ \mathbf{\bar{v} = 0.01},\ \sigma = \mathbf{0.5},\ \rho = \mathbf{0.5}$.

Pour le pricing d'une telle option, la seule méthode (envisageable) de pricing et/ou de calcul de couverture est une méthode de Monte Carlo ou de Quasi-Monte Carlo. Ici la direction optimale d'échantillonnage est un vecteur de $\mathbb{R}^{2d}$ car pour simuler une trajectoire du sous-jacent, l'on doit simuler aussi une trajectoire de la volatilité. Comme dans les exemples précédents, on voit que la procédure réduit de façon significative la variance Monte Carlo de l'estimation.

Tableau 4 Réduction de Variance estimée pour un Call Asiatique sous le modèle à volatilité stochastique de Heston

		IS				Brute	
$\mathbf{v_0}$	$\boldsymbol{\alpha}$	$\mathbf{K/S_0}$	**Price**	**Err**	**VR**	**Price**	**Err**
	1×10^3	1.3	0.35	0.004	16	0.38	0.02
	1×10^3	1.2	0.85	0.005	17	0.90	0.02
0.04	5×10^2	1.1	2.22	0.005	10.4	2.32	0.03
	8×10^1	1.0	6.12	0.01	6.4	6.25	0.04
	2×10^1	0.9	13.71	0.02	5.6	13.87	0.05
	1×10^1	0.8	22.60	0.02	5.5	22.81	0.05
	8×10^2	1.3	0.93	0.006	22	0.99	0.03
	4×10^2	1.2	1.84	0.009	16	1.93	0.04
0.09	2×10^2	1.1	3.73	0.01	11	3.88	0.05
	1×10^2	1.0	7.55	0.02	8.1	7.75	0.06
	2×10^1	0.9	14.15	0.03	7.2	14.37	0.07
	1×10^1	0.8	22.65	0.03	7.6	22.89	0.07

Nous avons utilisé 20,000 Tirages Monte Carlo et $\mathbf{n = 50}$ pas de discretisations. Les paramètres du Call sont $\mathbf{S_0 = 100}$, $\mathbf{r = 0.1}$, $\mathbf{T = 1, k = 2}$, $\mathbf{a = 0.01}$, $\sigma = \mathbf{0.5}$, $\rho = \mathbf{0.5}$.

3.5.3 Le modèle de Cox Ingersoll et Ross

Notre dernier exemple en finance concerne le pricing d'un Cap sur taux d'interêt dans le modèle de Cox Ingersoll et Ross (CIR 1985) où le taux suit la dynamique suivante:

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

avec W un mouvement Brownien standard et k, a, σ des constantes réelles positives. Il a été prouvé dans [ROG95] que lorsque $d \equiv 4ak/\sigma$ est un entier, le processus $(r_t)_{t\geq 0}$ suit la même loi que le processus $(\|X_t\|^2)_{t\geq 0}$, X étant le processus d-dimensionel solution de l'EDS

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

avec comme condition initiale X_0, le vecteur de taille d dont toutes les composantes sont égales à $\sqrt{r_0 d}$.

Le processus X est un processus d'*Orstein Uhlenbeck* dont la loi exacte est connue. On montre sans difficulté que sur une grille de points $t_j = jh$, $h > 0$ avec $j = 0, ...,$ la i^{eme} composante de ce processus est simulée de façon exacte par

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

où $G_1, ..., G_{dn}$ sont des réalisations indépendantes de la loi de Gauss centrée réduite. Nous nous plaçons dans le cas où $d = 1$, afin de séparer l'étude de l'erreur statistique dûe à la variance, que nous faisons, de l'erreur dûe au biais éventuels provenant de la discrétisation.

Tableau 5 Réduction de Variance estimée pour un Cap de taux d'intérêt sous le modèle CIR

		IS				**Brute**	
T	$\boldsymbol{\alpha}$	**K**	**Price**	**Err**	**VR**	**Price**	**Err**
	1×10^{-3}	0.044	30.215	0.006	33	30.258	0.035
	1×10^{-3}	0.054	15.311	0.008	18	15.344	0.033
0.25	1×10^{-1}	0.064	3.565	0.006	11	3.579	0.021
	1×10^{1}	0.074	0.402	0.001	24	0.405	0.006
	2×10^{2}	0.084	0.029	0.0001	103	0.031	0.001
	1×10^{-3}	0.044	28.631	0.009	25	28.682	0.046
	1×10^{-3}	0.054	14.976	0.01	16	15.015	0.041
0.50	1×10^{-2}	0.064	4.622	0.008	11	4.640	0.027
	1×10^{-1}	0.074	1.024	0.003	17	1.028	0.012
	1×10^{1}	0.084	0.190	0.001	44	0.193	0.005

Nous avons utilisé 50,000 Tirages Monte Carlo et $\mathbf{n = 16}$ dates de fixing. Les paramètres du Cap sont $\mathbf{r_0 = 0.064}$, $\mathbf{k = 0.05}$, $\mathbf{d = 1}$, $\sigma = \mathbf{0.08}$, $\rho = \mathbf{0.5}$, pour un nominal de **100**.

Soit un Cap de taux d'intérêt, de maturié T, de taux strike K, sur un nominal de M. On suppose que les dates de paiement sont $t_1 < t_2 < ... < t_n$ et $t_{n+1} = T$. A chaque date de paiement t_{i+1}, $i = 0, 1, ..., n$ l'échange de flux de ce Cap est $M(r_{t_i} - K)_+$. Le payoff de ce Cap est la somme de tous les flux actualisés et vaut

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

Comme le note Glasserman, Heidelberger et Shahabuddin [GHS99a], cette formulation ne tient pas vraiment compte de la distinction entre "l'univers continu" et "le monde discrèt" en termes d'actualisation (voir [MR97] ou [HUL03]). Néanmoins, cet exemple reste instructif sur l'efficacité de la procédure de réduction de variance que nous proposons. Les tests sont résumés dans le Tableau 5. On y voit que le taux de réduction de variance est important et varie de 10 à 100. Ce taux de réduction est peu influencé par le paramètre[2] α utilisé pour définir la suite de pas de convergence.

3.5.4 L'exemple de la fiabilité

Nos terminons les essais numériques sur un problème simple de fiabilité: l'estimation Monte Carlo de la durée de vie moyenne **MTTF** d'un système. Nous supposons pour cela que le temps de défaillance T du système suit une loi de Weibull $Wei(\beta, \sigma, \gamma)$ dont la fonction de répartition est donnée par

$$F(t) = 1 - \exp\left(-(\tfrac{t-\gamma}{\sigma})^{\beta}\right), \quad t > \max(0, \gamma),$$

où $\gamma \in \mathbb{R}$, et β, $\sigma > 0$ sont des constantes. Il est facile de vérifier que

$$\mathbf{MTTF} = \gamma + \sigma\Gamma(\tfrac{1+\beta}{\beta}),$$

avec $\Gamma(\cdot)$ la fonction Gamma. Par suite dans ce cas particulier, le calcul de **MTTF** se fera par l'utilisation d'une approximation numérique de la fonction $\Gamma(\cdot)$. Nous avons implémenté cette technique naturelle et dans le tableau 6, CF désigne les valeurs obtenues dans ce cas. Pour mettre en œuvre notre méthode, nous observons d'abord que la fonction de hasard (le taux de défaillance cumulé) du système est donnée par

$$\Lambda(t) = \left(\frac{t-\gamma}{\sigma}\right)^{\beta}, \quad t > \max(0, \gamma),$$

de sorte que son inverse ψ dont nous avons besoin est connue explicitement et vaut

$$\psi(x) = \sigma x^{\frac{1}{\beta}} + \gamma.$$

L'on peut noter dans ce cas que l'hypothèse (H_{14}) est facilement vérifiée. Le taux de réduction de variance obtenue est particulièrement intéressant lorsque β est petit. C'est ce cas qui correspond aux grandes valeurs de **MTTF**. On peut noter au passage l'apparition d'un biais important dans l'estimation Monte Carlo standard du **MTTF** lorsque le paramètre β est de plus en plus petit, et que ce biais est automatiquement corrigé par notre méthode.

[2]Il faut remarquer dans le Tableau 5 les variations des valeurs du paramètre α de la suite de pas de convergence. Ici le payoff n'est plus renormalisé comme dans les cas précédents. Nous y reviendrons à la fin de ce chapitre

Tableau 6 Réduction de Variance estimée pour le temps moyen de survie **MTTF** d'un système, sous la loi de Weibull

		IS			**Brute**		**CF**
β	α	**MTTF**	**Err**	**VR**	**MTTF**	**Err**	**MTTF**
0.2	0.0001	119.671	1.641	68	104.448	13.512	120
0.25	0.1	24.059	0.331	37	22.528	2.001	24
0.5	0.1	1.999	0.018	12	2.011	0.061	2
0.75	0.5	1.190	0.009	7.0	1.203	0.023	1.191
1	1	0.999	0.006	5.4	1.010	0.014	1
1.5	1	0.902	0.004	4.2	0.909	0.009	0.903
2	1	0.886	0.003	3.7	0.891	0.007	0.886

Les résultats sont basés sur 5,000 tirages Monte Carlo, et les paramètre de la loi sont $\theta_0 = \mathbf{1.0}$, $\sigma = \mathbf{1.0}$, $\gamma = \mathbf{0.0}$.

3.6 Quelques remarques pratiques

Notre but ici est d'exposer quelques indications intéressantes pour une bonne implémentation d'un algorithme stochastique de type Robbins–Monro.

- Le choix de $(\gamma_{\mathbf{n}})_{\mathbf{n}\geq\mathbf{0}}$

Nous rappelons que les algorithmes stochastiques que nous avons vus jusqu'ici ont été utilisés pour estimer asymptôtiquement le minimum θ^* d'une certaine fonction V défini par

$$V(\theta) = \mathbb{E}[H(\theta, X)].$$

L'observation de cet algorithme (*cf* Chapitre 2) notée $(Y_n)_{n\geq 1}$ vérifie

$$\mathbb{E}[Y_{n+1}/\mathcal{F}_n] = \nabla V(\theta_n),$$

avec $\mathcal{F}_n = \sigma\{\theta_0, X_k, 0 \leq k \leq n\}$ la filtration propre de l'algorithme. Soit $\nabla^2 V(\theta^*)$ la matrice hessienne (lorsqu'elle est bien définie) de V prise au point θ^* (nous avons montrer dans le Chapitre 2 que cette matrice est bien définie). Il est prouvée (par exemple dans [PEL96]) que si L désigne la plus grande valeur propre de $\nabla^2 V(\theta^*)$, alors les meilleures suites de pas de convergence $(\gamma_n)_{n\geq 0}$ parmi toutes les suites possibles sont données par

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

Par "meilleures", nous entendons toutes les suites qui permettent d'obtenir la vitesse optimale (en $1/\sqrt{n}$) de convergence en loi de l'algorithme $(\theta_n)_{n\geq 0}$.

Malheureusement, le calcul pratique de la valeur propre L est impossible, puisque θ^* est inconnu. Mais cela nous suggère l'heuristique suivante:

- Pour "pricer" un Call ou un Put, avec notre méthode de Monte Carlo adaptative, il peut être plus intéressant de normaliser la fonction de payoff ϕ. En effet dans ce cas la fonction à minimiser est donnée par

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

et si l'on pose $\hat{\varphi}(x) \stackrel{\Delta}{=} \varphi(x)/S_0$, avec S_0 le prix spot du sous-jacent, alors il est clair que

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

où

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

L'idée consiste, alors, à mettre en œuvre un algorithme de Robbins–Monro pour $\hat{V}$ plutôt que pour V. On vérifie qu'une bonne valeur de α obtenue pour une option de spot et strike S_0^1 et K^1 respectivement, devrait donner de bons résultats pour le pricing d'une autre option de spot S_0^2 et de strike K^2 pourvu que $S_0^1/K^1 = S_0^2/K^2$. Dans tous nos tests numériques, nous avons pris en considération cette remarque sauf pour le pricing des Caps de taux d'intérêt (la valeur initiale du taux étant déjà ¡ 1).

- Si les observations Y_{n+1} étaient déterministes, alors l'algorithme de Robbins–Monro à pas décroissant ou constant serait réduit à un schéma d'Euler d'une équation différentielle ordinaire. Cette simple remarque permet de comprendre que la variabilité des algorithmes stochastiques peut être controlée par la variance conditionnelle de Y_{n+1}. On y revient un peu plus en détail à la section suivante.

- Le choix de θ_0

Le choix de la valeur initiale θ_0 de l'algorithme $(\theta_n)_{n\geq 0}$ peut s'avérer important. En général, dans le cas déterministe, il existe des méthodes empiriques pour un tel choix, alors que dans le cas stochastique, à cause de la présence de l'aléa, l'importance de ce choix est moins significative. Dans le cas particulier de notre méthode MCA, du fait de la prise en compte itérative des valeurs de $(\theta_n)_{n\geq 0}$ dans la moyennisation Monte Carlo, il importe de bien choisir la valeur de θ_0. Mais en générale l'on ne dispose pas de moyens efficaces pour le faire. Une manière très simple de contourner le problème est de considérer (dans la moyennisation Monte Carlo) les valeurs de l'algorithme $(\theta_n)_{n\geq 0}$ à partir d'un certain rang n_0. Dans ce cas $(\star)_a$ devient

$$\frac{1}{n-n_0}\sum_{k=n_0+1}^{n} \varphi(X_k+\theta_{k-1})e^{-\theta_{k-1}\cdot X_k+\|\theta_{k-1}\|^2},$$

alors que $(\star)_b$ est remplacé par

$$\frac{1}{n-n_0}\sum_{k=n_0+1}^{n}\frac{1}{\theta_{k-1}}\psi\left(\frac{X_k}{\theta_{k-1}}\right)e^{X_k-\frac{X_k}{\theta_{k-1}}}.$$

Dans nos tests, nous avons utilisé $n_0 = 500$ dans le premier cas et $n_0 = 100$ dans le second.

3.7 Méthodes de réduction de variance des Algorithmes Stochastiques

Plusieurs pistes peuvent être explorées pour l'amélioration de la convergence numérique (liée à la stabilité de l'algorithme stochastique utilisé) de notre méthode de réduction de variance. Sans être exhaustifs, nous allons tenter dans cette section d'y apporter quelques éléments de réponse.
Rappelons la forme générale des algorithmes stochastiques de type Robbins-Monro

$$\theta_{n+1} = \theta_n - \gamma_{n+1}h(\theta_n) + \gamma_{n+1}(\varepsilon_{n+1} + r_{n+1}), \quad n \geq 0, \tag{.3.13}$$

avec $\mathbb{E}[\varepsilon_{n+1}/\mathcal{F}_n] = 0$, $(\mathcal{F}_n)_{n\geq 0}$ étant la filtration naturelle de $(\theta_n)_{n\geq 0}$ et $(r_n)_{n\geq 1}$ un terme résiduel supplémentaire, le plus souvent nul, qui peut provenir par exemple d'une projection de l'algorithme stochastique. Nous avons vu dans le Chapitre 1, qu'un tel algorithme pouvait être vu comme un schéma d'Euler de pas décroissant $(\gamma_n)_{n\geq 0}$ de l'équation différentielle ordinaire associée à la fonction $(-h)$, à savoir

$$\dot{\theta}(t) = -h(\theta(t)) \quad t \geq 0, \tag{ODE_h}$$

auquel on a ajouté le bruit $(\varepsilon_n + r_n)_{n\geq 1}$. La convergence de cet algorithme repose fondamentalement sur le controle du bruit $(\varepsilon_n + r_n)$ (voir les Théorèmes .5 et .6). Pour plus de simplicité, nous allons supposer dans la suite que le terme $(r_n)_{n\geq 1}$ est nul. Il est clair que la suite $(\theta_n)_{n\geq 0}$ convergerait d'autant plus vite que le bruit $(\varepsilon_n)_{n\geq 1}$ serait de taille "petite". En effet si ce bruit était nul, alors l'algorithme $(\theta_n)_{n\geq 0}$ se reduirait à un schéma d'Euler *déterministe* de l'équation différentielle ODE_h. L'accélération de la vitesse de convergence (numérique) de l'algorithme stochastique repose donc en grande partie sur la réduction de sa variabilité (*i.e.* l'atténuation de son caractère stochastique). Nous commencerons par spécifier ce que nous entendons par variabilité des algorithmes stochastiques. Ensuite dans le cas particulier des algorithmes à dynamique gaussienne, nous essayerons de trouver de nouvelles représentations des observations dont nous comparerons les variabilités.

Réduire la variabilité de l'algorithme stochastique $(\theta_n)_{n\geq 0}$ consiste à rechercher une nouvelle représentation du bruit $(\varepsilon_n)_{n\geq 1}$, que nous noterons $(\hat{\varepsilon}_n)_{n\geq 0}$ telle que

$$\mathbb{E}[\hat{\varepsilon}_{n+1}/\mathcal{F}_n] = 0 \quad \text{et} \quad \mathbb{E}[\|\hat{\varepsilon}_{n+1}\|^2/\mathcal{F}_n] < \mathbb{E}[\|\varepsilon_{n+1}\|^2/\mathcal{F}_n], \quad \forall\, n \geq 0. \tag{.3.14}$$

Ces conditions sont les analogues des conditions (.2.13) qui portent sur les observations de l'algorithme. On comprend bien qu'il est quasi-impossible de faire une étude générale de ce type de problème, sans spécifier explicitement la forme de ε_n. On rappelle que l'observation $F(\theta_n, X_{n+1})$ est liée au bruit ε_{n+1} par la relation

$$\varepsilon_{n+1} = h(\theta_n) - F(\theta_n, X_{n+1}), \quad n \geq 0.$$

Pour des raisons de faisabilité, nous allons remplacer les conditions (.3.14) par les conditions suivantes: trouver une représentation $\hat{F}(\theta, x)$ telle que

$$\mathbb{E}[\hat{F}(\theta_n, X_{n+1})/\mathcal{F}_n] = h(\theta_n) \quad p.s. \quad \text{et} \quad \frac{1}{N}\sum_{n=0}^{N-1}\left\|\hat{F}(\theta_n, X_{n+1})\right\|^2 << \frac{1}{N}\sum_{n=0}^{N-1}\|F(\theta_n, X_{n+1})\|^2, \tag{.3.15}$$

D'après le Théorème .13 on devrait avoir,

$$\frac{1}{N}\sum_{n=0}^{N-1}\left\|\hat{F}(\theta_n, X_{n+1})\right\|^2 \underset{N\to+\infty}{\overset{p.s.}{\longrightarrow}} \mathbb{E}\left[\|F(\theta^*, X)\|^2\right].$$

De plus comme

$$\frac{1}{N}\sum_{n=0}^{N-1}\left\|\hat{F}(\theta_n, X_{n+1})\right\|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{\gamma_{n+1}^2}\|\theta_{n+1} - \theta_n\|^2,$$

on peut raisonnablement considérer la quantité

$$VAR := \frac{1}{N}\sum_{n=0}^{N-1}\left\|\hat{F}(\theta_n, X_{n+1})\right\|^2$$

comme une "bonne" mesure de variabilité globale de l'algorithme (.3.13). Nous limiterons nos investigations numériques au cadre gaussien considéré jusqu'à maintenant. Ce cadre se caractérise par des observations définies en (.2.10) par

$$F(\theta_n, X_{n+1}) = (\theta_n - X_{n+1})\varphi^2(X_{n+1})e^{-\theta_n\cdot X_{n+1} + \frac{1}{2}\|\theta_n\|^2}, \quad n \geq 0.$$

3.7.1 Méthode des Variables Antithétiques et Fonction d'Importance pour les observations

Comme X est un vecteur gaussien on peut appliquer le théorème de Girsanov au vecteur aléatoire $F(\theta, X) = (\theta - X)\varphi^2(X)e^{-\theta\cdot X+\frac{1}{2}\|\theta\|^2}$, pour shifter la loi de X. Cela donne

$$\mathbb{E}[F(\theta, X)] = \mathbb{E}[F(\theta, X+\theta)e^{-\theta\cdot X-\frac{1}{2}\|\theta\|^2}], \quad \forall\, \theta \in \mathbb{R}^d, \tag{.3.16}$$

$$= \mathbb{E}[F(\theta, X-\theta)e^{+\theta\cdot X-\frac{1}{2}\|\theta\|^2}], \quad \forall\, \theta \in \mathbb{R}^d. \tag{.3.17}$$

En posant $F_1(\theta, x) = F(\theta, x+\theta)e^{-\theta\cdot x-\frac{1}{2}\|\theta\|^2}$ et $F_2(\theta, x) = F(\theta, x-\theta)e^{\theta\cdot x-\frac{1}{2}\|\theta\|^2}$ on montre par de simples manipulations que

$$F_1(\theta, X) = -X\varphi^2(X+\theta)e^{-2\theta\cdot X-\|\theta\|^2}, \quad \theta \in \mathbb{R}^d$$

et

$$F_2(\theta, X) = (2\theta - X)\varphi^2(X-\theta)e^{\|\theta\|^2} \quad \theta \in \mathbb{R}^d.$$

Pour ces représentations on a encore

$$\mathbb{E}[F_1(\theta_n, X_{n+1})/\mathcal{F}_n] = \mathbb{E}[F_2(\theta_n, X_{n+1})/\mathcal{F}_n] = h(\theta_n) \quad p.s.$$

Lorsque $\mathbb{P}(\varphi(X) \simeq 0) \simeq 1$, il peut être plus intéressant d'utiliser F_1 ou F_2 (ou une combinaison convexe des deux) comme observation plutôt que la représentation F.
La méthode des variables antithétiques consiste à considérer comme observation

$$F_a(\theta, X) = \frac{1}{2}\left(F(\theta, X) + F(\theta, -X)\right). \tag{.3.18}$$

X étant un vecteur gaussien centré, on a encore

$$\mathbb{E}[F_a(\theta_n, X_{n+1})/\mathcal{F}_n] = h(\theta_n) \quad p.s.$$

Nous nous sommes contentés de l'exemple du Call sous le modèle de Black et Scholes pour effectuer les tests de variabilité de l'algorithme (.1.11-.1.13). Les résultats sont regroupés dans les Tableaux 7 et 8 ci-dessous. Ces résultats concernent le cas d'un Call Européen avec d'abord une faible volatilité ($\sigma = 10\%$) puis une volatilité élevée ($\sigma = 30\%$). Les quatre dernières colonnes de ces tableaux contiennent le cœfficient de variabilité VAR défini plus haut pour l'algorithme stochastique (.1.11-.1.13) avec respectivement $F(\theta_n, X_{n+1})$, $F_a(\theta_n, X_{n+1})$, $F_1(\theta_n, X_{n+1})$ et $F_2(\theta_n, X_{n+1})$ pour observations. En dessous du coefficient de variabilité, apparait entre parenthèses la valeur estimée du drift optimal θ^* par l'algorithme correspondant. En fait nous avons

délibérément "oublié" les premières itérations des algorithmes dans l'estimation de la variabilité, car nous avons observé que ces premières itérations prenaient souvent des valeurs excessives dûes probablement à des valeurs de γ_n trop grandes au début. Ainsi le vrai coefficient de variablité que nous avons estimé est

$$VAR := \frac{1}{N-n_0}\sum_{n=n_0}^{N-1}\left\|\hat{F}(\theta_n, X_{n+1})\right\|^2,$$

avec $n_0 = 100$ dans nos tests. Pour faire ces tests nous avons utilisé les mêmes valeurs du paramètre α de la suite de pas de convergence de l'algorithme que celles qu'on a utilisé dans le Chapitre 2 pour le même type d'options.
On voit clairement dans les Tableaux 7 et 8 que l'observation F_1 est, globalement, la plus satisfaisante dans tous les cas, dans la mesure où elle fournit souvent la variabilité la plus faible.

Tableau 7 Variabilité de l'algorithme (.1.11-.1.13) pour un Call Européen peu volatile

α	K/S_0	VAR_s	VAR_a	VAR_1	VAR_2
10	0.6	0.03362 (0.229)	0.00356 (0.234)	0.03036 (0.228)	0.04504 (0.228)
	0.8	0.00294 (0.407)	0.00094 (0.413)	0.00225 (0.406)	0.00629 (0.409)
20	1.0	0.00005 0.775	0.00003 0.747	0.00002 (0.779)	0.00043 (0.780)
100	1.2	0.00001 (0.437)	0.00001 (0.412)	0.00000 (0.436)	0.00004 (0.439)

Nous avons utilisé RM=10,000 trajectoires simulées pour la convergence de l'algorithme. Les paramètres du call sont: $\mathbf{r = 5\%}$, $\sigma = \mathbf{10\%}$, $\mathbf{T = 1.0}$.

Tableau 8 Variabilté de l'agorithme (.1.11-.1.13) pour un Call Européen volatile

α	K/S_0	VAR_s	VAR_a	VAR_1	VAR_2
25	0.6	0.03837 (0.656)	0.02037 (0.666)	0.02053 (0.675)	0.18557 (0.677)
	0.8	0.00572 (0.909)	0.00332 (0.919)	0.00239 (0.924)	0.07911 (0.913)
	1.0	0.00061 (1.231)	0.00032 (1.234)	0.00019 (1.230)	0.02897 (1.270)
50	1.2	0.00006 (1.561)	0.00003 (1.539)	0.00001 (1.562)	0.00796 (1.580)
	1.4	0.00004 (1.443)	0.00003 (1.388)	0.00000 (1.397)	0.00361 (1.441)
	1.6	0.00005 (1.145)	0.00005 (1.023)	0.00000 (1.142)	0.00221 (1.098)

Nous avons utilisé RM=10,000 trajectoires simulées pour la convergence de l'algorithme. Les paramètres du call sont: $\mathbf{r = 5\%}$, $\sigma = \mathbf{30\%}$, $\mathbf{T = 1.0}$.

Remarque .36. Pour finir, nous évoquons deux techniques (générales) de moyennisation, de l'algorithme $(\theta_n)_{n\geq 0}$. La première qui consiste à moyenniser les itérations de l'algorithme est la méthode de Ruppert et Polyak, alors que la deuxième consiste à moyennisation les observations de l'algorithme. On peut montrer que, théoriquement, ces techniques sont optimales (en un sens que nous ne détaillons pas, pour plus de précision à ce sujet voir [KY03], [POL90] ou [RUP88]). Toutefois sur le plan strictement numérique, il est bien connu que la moyennisation ralentit considérablement la convergence (numérique) de l'algorithme. Lorsqu'on sait montrer que l'algorithme $(\theta_n)_{n\geq 0}$ converge $(p.s.)$ vers une valeur θ^*, on peut être tenter de dire que l'algorithme moyennisé suivant

$$\bar{\theta}_n = \frac{1}{N_n} \sum_{k=n-N_n+1}^{k=n} \theta_k \tag{.3.19}$$

est meilleur estimateur de θ^* que θ_n ne l'est, et ce pour une suite $(N_n)_n$ qui tend vers l'infini et $n - N_n \geq 0$. En réalité cela dépend beaucoup de la forme de la suite γ_n de pas de convergence de l'algorithme θ_n. Polyak et Ruppert ont montré que si γ_n décroissait

moins vite que $1/n$ alors la remarque ci-dessus est vraie: (.3.19) est préférable à θ_n pour estimer θ^*. On peut se référer à [KY03] (chap.11), ou à [DUF96] (chap.4.III) pour plus de détails.

Remarque .37. La moyennisation des observations conduit à l'algorithme suivant:

$$\theta_{n+1} = \hat{\theta_n} - (n+1)\gamma_{n+1}\hat{F}(\theta_n, X_{n+1}), \tag{.3.20}$$

avec

$$\hat{F}(\theta_n, X_{n+1}) = \frac{1}{n}\sum_{k=1}^{n}\hat{F}(\theta_k, X_{k+1}), \quad \text{et} \quad \hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n}\hat{\theta}_k. \tag{.3.21}$$

De nouveau, sous de bonnes hypothèses, on peut montrer que cet algorithme est asymptôtiquement "optimal" (voir [KY03] (chap.11) et [DUF96] (chap.2.I) et les nombreuses références qui s'y trouvent).

Même si ces deux techniques de moyennisation permettent de contourner le problème difficile du choix de la suite de pas de convergence γ_n, Dans nos implémentations, elles n'ont pas donné satisfaction. En effet, la seule moyennisation peut multiplier par un facteur de 10 (et même plus!) le temps global d'exécution de la procédure de Monte Carlo avec réduction de variance par un algorithme stochastique non moyennisé. Dans les procédures de réductions de variance que nous avons cherchées à développer, nous avons porté une attention particulère à l'obtention d'une bonne réduction de variance pour un coût de calcul "raisonnable". Les techniques de moyennisation nous ont paru à cet effet trop chères.

3.8 Réduction de Variance et Algorithmes Stochastiques à gain constant

Jusqu'ici nous nous sommes servi des algorithmes stochastiques à pas décroissants pour développer nos méthodes de réduction de variance. L'objectif de cette section est de montrer que, grâce à leurs propriétés ergodiques, les algorithmes stochastiques à pas constants permettent également d'obtenir des réductions de variance intéressantes.

3.8.1 Cadre gaussien

Revenons à l'objectif initial de l'estimation du minimiseur de $\mathbb{V}\mathrm{ar}(g(\theta, X))$ avec X un vecteur gaussien de taille d et $g(\theta, x) = \varphi(x+\theta)e^{-\theta\cdot x - \frac{1}{2}\|\theta\|^2}$.
On a vu (Proposition 2.3.1 et remarque .18) que ce problème était équivalent à

$$h(\theta) = 0, \quad \theta \in \mathbb{R}^d, \tag{.3.22}$$

avec $h(\theta) = \mathbb{E}[F(\theta, X)]$ et $F(\theta, x) = (\theta - x)\varphi^2(x)e^{-\theta\cdot x + \frac{1}{2}\|\theta\|^2}$, dont la solution unique est notée $\theta^* \in \mathbb{R}^d$. On fixe $\theta_0 \in \mathbb{R}^d$ et on pose $V(\theta) = \frac{1}{2}\|\theta - \theta^*\|^2$. On considère $\lambda > 0$ et $R_\lambda > 0$ tels que

$$V(\theta_0) \leq \lambda \quad \text{et} \qquad R_\lambda^2/2 > \lambda + \|\theta^*\|^2 . \tag{.3.23}$$

Soit $K_\lambda = \left\{\theta, \in \mathbb{R}^d;\ \|\theta\|^2 \leq R_\lambda^2\right\}$ la boule fermée centrée de $\mathbb{R}^d$ de rayon R_λ.

Remarque .38. Il est facile de voir que $\theta_0 \in K_\lambda$, et donc d'après la Proposition 1.5.1 toute solution $(\theta(\cdot))$ de (EDO_h) de condition initiale θ_0 reste entièrement dans l'interieur du compact K_λ. On considèrera ce K_λ comme compact de projection de l'algorithme stochastique (.1.20).

En pratique, on définit l'algorithme stochastique *projeté* $(\theta_n^\gamma)_{n\geq 0}$ de façon itérative en posant: $\theta_0^\gamma = \theta_0$ et pour $n \geq 0$

$$\theta_{n+1}^\gamma = \begin{cases} \theta_n^\gamma - \gamma_{n+1}Y_{n+1}^\gamma & \text{si} \quad \|\theta_n^\gamma - \gamma_{n+1}Y_{n+1}^\gamma\| \leq R_\lambda, \\ \theta_n^\gamma & \text{sinon} \end{cases} \tag{.3.24}$$

avec $Y_{n+1}^\gamma = F(\theta_n^\gamma, X_{n+1})$ et $(X_n)_{n\geq 1}$ une suite *i.i.d.* de même loi que X.

Remarque .39. Ici la filtration propre de l'algorithme ne dépend pas de γ et est donnée par $\mathcal{F}_n = \sigma\{X_k,\ k \leq n\}$. Il est clair qu'on peut réecrire cet algorithme sous la forme

$$\theta_{n+1}^\gamma = \Pi_{K_\lambda}(\theta_n^\gamma - \gamma Y_{n+1}^\gamma),\ n \geq 0, \quad \theta_0^\gamma \in K_\lambda.$$

par définition même puisque $\theta_0 \in K_\lambda$.

On a le résultat suivant

Théorème .16. *Si $\mathbb{E}[\varphi^{2a}(X)] < +\infty$ pour un $1 < a \leq 2$, alors les hypothèses (H_7) à (H_{10}) sont vérifiées et le processus interpolé $(\theta^\gamma(t))_{t\geq 0}$ associé à l'algorithme stochastique* (.3.24) *converge en loi vers la solution de l'EDO*

$$\dot{\theta}(t) = -h(\theta(t)), \quad h(\theta) = \mathbb{E}[F(\theta, X)]$$

de condition initiale $\theta(0) = \theta_0 \in K_\lambda$.

Preuve. D'abord remarquons que les hypothèses (H_8), (H_9) et (H_{10}) sont trivialement satisfaites en prenant

$$h_n^\gamma \equiv h \quad \text{et} \quad \bar{h} \equiv h.$$

Pour vérifier (H_7), choisissons $1 < b < a$. On a

$$\begin{aligned} \mathbb{E}\left[\left\|Y_{n+1}^\gamma\right\|^b\right] &= \mathbb{E}\left[\mathbb{E}\left[\left\|Y_{n+1}^\gamma\right\|^b / \mathcal{F}_n\right]\right] \\ &= \mathbb{E}\left[s_b(\theta_n^\gamma)\right] \end{aligned}$$

avec

$$\begin{aligned} s_b(\theta) =& \mathbb{E}\left[\|\theta - X\|^b \varphi^{2b}(X) e^{-b\theta\cdot X + \frac{1}{2}\|\theta\|^2} \right] \\ \leq& \mathbb{E}\left[2^{b-1}(R_\lambda^b + \|X\|^b)\varphi^{2b}(X) e^{bR_\lambda \|X\| + \frac{b}{2}R_\lambda^2} \right] \\ <& +\infty. \end{aligned}$$

Ainsi $\exists L < +\infty$ tel que $\mathbb{E}\left[\left\| Y_{n+1}^\gamma \right\|^b \right] \leq L$ avec $b > 1$; l'hypothèse (H_7) est donc satisfaite et le Théorème .8 s'applique. Maintenant étudions le processus limite $(Z(t))_t$ du théorème. D'après le Théorème .8, il existe un processus intégrable $(z(t))_t$ tel que

$$z(t) \in -C(\theta(t)) \quad \text{et} \quad Z(t) = \int_0^t z(s)ds.$$

Or d'après la Remarque .38, $\forall\, t \geq 0$, $\theta(t) \in \overset{\circ}{K}_\lambda$. K_λ étant la boule fermée centrée de rayon R_λ choisi comme dans (.3.23), la Remarque .14 affirme que $\forall\, \theta \in \overset{\circ}{K}_\lambda$, $\quad C(\theta) = \{0\}$. Donc

$$\forall\, t \geq 0, \quad C(\theta(t)) = \{0\} \quad \text{et} \quad z(t) = 0.$$

Finalement comme $Z(0) = 0$ (en $t = 0$ il n'a y pas eu de projection) on a

$$Z(t) = 0, \quad \forall\, t \geq 0.$$

Ce qui prouve que toutes les sous-suites convergentes (faiblement) de $(\theta^\gamma(t))_t$ converge vers la même limite, en l'occurrence l'unique solution de EDO_h. Ce qui achève la démonstration du théorème. □

Remarque .40. En fait dans ce cadre gaussien qui est le notre, on peut vérifier que toutes les hypothèses du **Théorème 1.1** (page 319) de [KY03] sont satisfaites et que pour n assez grand

$$\frac{(\theta_n^\gamma - \theta(n\gamma))}{\sqrt{\gamma}} \xrightarrow[\gamma\to 0]{\mathcal{L}} \mathcal{N}(0,\, \Sigma),$$

avec $\Sigma = \mathbb{E}\left[F(\theta^*, X) F(\theta^*, X)^T \right]$.

3.8.2 Procédure de réduction de variance

Pour rendre notre discours plus cohérent et autonome, nous reprenons ici une partie de ce qui a été dit à la section 4 du Chapitre 2.
Notre objectif premier était de trouver une bonne méthode de réduction de variance pour l'estimation Monte Carlo de la quantité $\mathbb{E}[\varphi(X)]$ avec $X \sim \mathcal{N}(0,\, I_d)$.

Jusqu'ici, nous avons toujours estimé le paramètre θ^* (solution de (.3.22)) par un algorithme stochastique à pas décroissant. En effet, nous avons prouvé à chaque fois un résultat de convergence presque sûr de l'algorithme vers θ^*.
Pour les algorithmes à pas constant, il n'est plus possible de démontrer un résultat de convergence de ce type, mais un résultat plus faible de convergence en loi. Le Théorème .16 et la Remarque .40 affirment en effet que pour n assez grand et pour un pas γ suffisamment petit, l'algorithme (.3.24) vérifie:

$$\theta_n^\gamma \simeq \theta^* + \sqrt{\gamma}\, G,$$

où $G \sim \mathcal{N}(0, \Sigma)$. Heuristiquement, pour γ petit, le régime stationnaire de θ_n^γ se concentre autour de θ^*. C'est ce qui suggère d'essayer l'approximation Monte Carlo

$$\varphi(X) \simeq \frac{1}{N} \sum_{n=1}^{N} g(\hat{\theta}_n^\gamma, X_n), \qquad (ALGO_1)$$

avec $\hat{\theta}^\gamma$ une estimation asymptôtique de θ^* par l'algorithme θ_n^γ utilisant comme observation

$$Y_{n+1}^\gamma = (\theta_n^\gamma - X_{n+1})\varphi^2(X_{n+1}) e^{-\theta_n^\gamma \cdot X_{n+1} + \frac{1}{2}\|\theta_n^\gamma\|^2}.$$

La mise en œuvre de cette procédure de réduction de variance se fait exactement comme aux sections 4 et 5 du Chapitre 2 à la seule différence que l'algorithme (.1.11-.1.13) est remplacé par l'algorithme (.3.24) à pas constant (et petit).

Remarque .41. Le même procédé peut être mis en œuvre dans le cadre du prochain Chapitre. Il n'y a que le critère à minimiser qui change. Ce qui revient à résoudre l'équation (.4.21) ou (.4.20) par un algorithme stochastique à pas constant.

3.9 Illustration Numérique

Comme dans le Chapitre 2, nous mesurons l'efficacité de cette méthode de réduction de variance par le rapport de sa variance et de celle de la méthode de Monte Carlo sans réduction de variance. Dans le cas de l'approximation $(ALGO_1)$, ce ratio vaut

$$\mathbf{RV} = \frac{\frac{1}{N}\sum_{n=1}^{n=N} g^2(0, X_n) - \left(\sum_{n=1}^{n=N} g(0, X_n)\right)^2}{\frac{1}{N}\sum_{n=1}^{n=N} g^2(\hat{\theta}^\gamma, X_n) - \left(\sum_{n=1}^{n=N} g(\hat{\theta}^\gamma, X_n)\right)^2},$$

Quelques réglages s'imposent au niveau du choix effectif de la taille du compact K_λ de projection de l'algorithme (.3.24):
D'après la forme des observations de cet algorithme (surtout la partie en exponentielle

$e^{-\theta_n^\gamma \cdot X_{n+1} + \frac{1}{2}\left\|\theta_n^\gamma\right\|^2}$), il est facile de prédire que la norme $\|\theta^*\|$ de θ^* ne dépassera pas quelques unités. On prendra donc $R_\lambda = 5$ ou $R_\lambda = 10$ comme rayon du compact K_λ selon qu'on est en dimension grande ou pas. La Remarque .23 du Chapitre 1 sera prise en compte dans nos implémentations. Nous rappelons que d'après cette remarque il est plus commode de mettre en œuvre l'algorithme θ_n^γ en utilisant comme observations

$$\bar{Y}_{n+1}^\gamma = (\theta_n^\gamma - X_{n+1})\bar{\varphi}^2(X_{n+1})e^{-\theta_n^\gamma \cdot X_{n+1} + \frac{1}{2}\|\theta_n^\gamma\|^2}, \quad \text{avec} \quad \bar{\varphi}(x) = \varphi(x)/S_0.$$

Remarque .42. Toutes les valeurs de γ qui apparaitront dans la suite devrons être divisées par S_0^2 pour retrouver les "vraies petites valeurs" de γ.
Par exemple une valeur $\gamma = 2$ avec un spot $S_0 = 50$ vaut en réalité $\hat{\gamma} = 0.0008$

Nous avons effectué une série de tests qui confirment que les algorithmes à pas constants conduisent à des taux de réduction de variance du même ordre de grandeur que les algorithmes à pas décroissants. Nous présentons ici quelques uns seulement de ces tests numériques.
Les figures 1 et 2 montrent un exemple du type de dépendance qui existe entre le ratio de variance **RV** et le paramètre γ dans le cas simple d'un call européen à la monnaie sous le modèle de Black-Scholes. On y voit que la réduction de variance est importante pour un large intervalle de valeurs de γ et que cette réduction décroit légèrement lorsque γ devient de plus en plus grand. Ce qui est tout à fait normal puisqu'une grande valeur de γ correspond à un bruit important dans l'estimation de θ^* par θ_n^γ. En fait la Figure 2 est un zoom de la Figure 1 autour de la valeur $\gamma = 0.01$. Elle confirme le fait que γ devrait être petit mais pas trop, sinon les itérations de l'algorithme ne feront que "du sur-place". Dans la Tableau 1, sont résumés les résultats numériques obtenus pour un Call européen sur moyenne arithmétique. Rappelons que c'est l'un des cas d'options les plus simples pour lesquels le problème de pricing ne peut être abordé que par la simulation (à l'inverse du call asiatique sur moyenne continue). On peut ramarqer que le taux de réduction de variance obtenu est similaire à celui des tests numériques du Chapitres 2 et des résultats de [GHS99a] (Tableau 5.1). Le léger biais qui apparait entre nos prix et ceux de [GHS99a] est dû au fait qu'on a utilisé au total beaucoup moins de simulations Monte Carlo.
Un autre cas d'option simple dont le problème de pricing requiert la simulation est celui des options sur panier d'actions (ou sur indice). de payoff

$$\max\left(\sum_{i=1}^{d} a_i S_T^i - K, 0\right),$$

où les a_i sont des poids positifs vérifiant $\sum_{i=1}^{d} a_i = 1$.

Figure 3.1: continum du ratio de variance **RV** par rapport au pas de convergence γ pour un Call européen à la monnaie. $\mathbf{S_0 = K = 50}$, $\mathbf{r = 0.05}$, $\mathbf{sigma = 0.3}$, $\mathbf{T = 1}$, $\mathbf{MC = 40,000}$, $\mathbf{RM = 10,000}$.

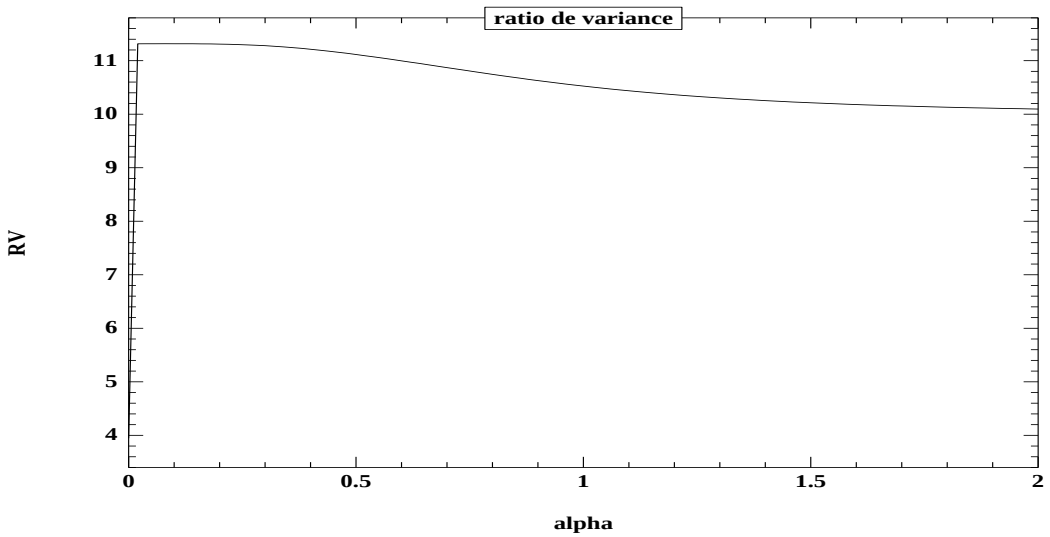
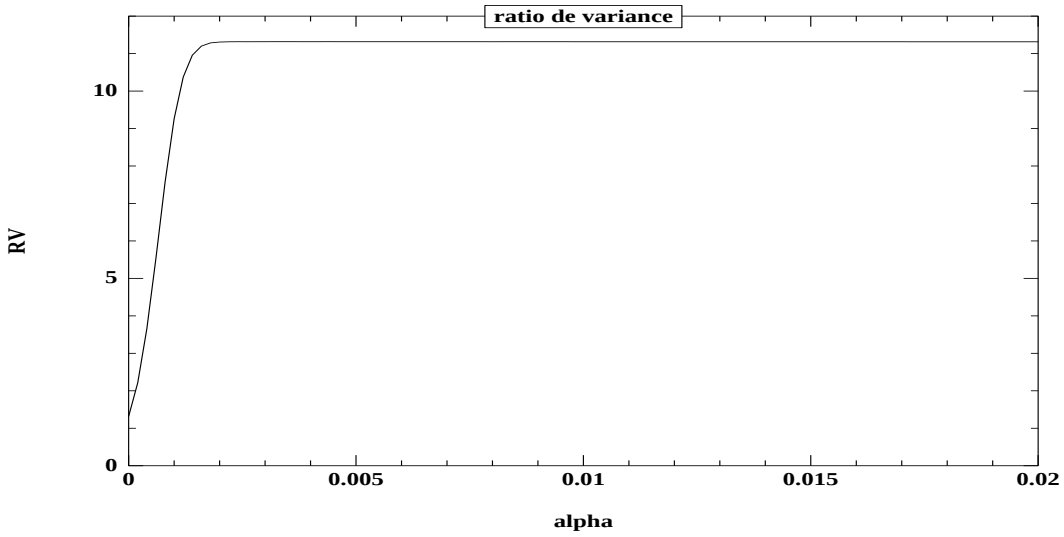


Figure 3.2: continum du ratio de variance **RV** par rapport au pas de convergence γ pour un Call européen à la monnaie. $\mathbf{S_0 = K = 50}$, $\mathbf{r = 0.05}$, $\mathbf{sigma = 0.3}$, $\mathbf{T = 1}$ $\mathbf{MC = 40,000}$ $\mathbf{RM = 10,000}$.



Sous la probabilité de pricing (probabilité risque-neutre) l'actif S_t^i suit la dynamique de Black–Scholes

$$\frac{dS_t^i}{S_t^i} = rdt + \sum_{i=1}^{d} \sigma_{ij} dW_t^i, \quad S_0^i > 0$$

avec $W = (W^1, \ldots, W^d)$ un mouvement brownien $d-$dimensionnel et $\sigma = (\sigma_{ij})_{ij}$ la matrice de volatilité de marché. Par souci de simplicité de notre exposé, nous présentons dans le Tableau 2 le cas où $\sigma_{ij} = \sigma > 0$ et $S_0^i = S_0$. On y voit par exemple que les valeurs de γ sont plutôt petites, à cause du problème bien connu de la forte variance des options sur panier. D'ailleurs la réduction de variance obtenue est plus importante que dans le cas asiatique ci-dessus.

Tableau 1 Réduction de Variance estimée pour Call Asiatique sous le modèle de Black-Scoles

	Paramètres			**IS**	
n	$\boldsymbol{\gamma}$	$\boldsymbol{\sigma}$	$\mathbf{K/S_0}$	**PrixRM**	$\mathbf{RV_2}$
16	0.01	0.1	0.8	10.80	22.1
	0.5		1.0	1.93	7.3
	500		1.2	0.006	142.7
	0.01	0.3	0.8	11.09	9.7
	0.05		1.0	4.19	9.7
	0.5		1.2	1.09	19.1
64	0.005	0.1	0.8	10.73	11.4
	0.3		1.0	1.83	7.3
	10		1.1	0.17	20.7
	0.01	0.3	0.8	10.94	9.2
	0.05		1.0	3.99	9.6
	0.5		1.2	0.98	18.7

Les résultats sont basés sur 10,000 tirages Monte Carlo et 4,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

Pour finir, nous anticipons un peu en testant l'estimation

$$\varphi(X) \simeq \frac{1}{N} \sum_{n=1}^{N} g(\bar{\theta}_{n-1}^{\gamma}, X_n), \qquad (ALGO_2)$$

où $\bar{\theta}_n^\gamma$ est l'algorithme (.3.24) avec simplement comme observations (voir (.4.21) du Chapitre 4)

$$Y_{n+1}^\gamma = \varphi(X_{n+1})(\theta_n^\gamma - X_{n+1}).$$

Tableau 2 Réduction de Variance estimée pour un Call Basket Européen sous le modèle de Black-Scoles

	Paramètres			**IS**	
n	$\boldsymbol{\gamma}$	$\boldsymbol{\sigma}$	$\mathbf{K/S_0}$	**PrixRM**	**RV**
10	0.001	0.3	0.6	27.22	37.2
	0.001		1.0	19.02	35.7
	0.001		1.4	13.83	39.6
	0.001	0.1	0.6	21.62	15.5
	0.01		1.0	7.41	11.6
	0.1		1.4	1.80	24.2
20	0.0005	0.3	0.6	31.77	103.9
	0.001		1.0	25.61	96.1
	0.001		1.4	21.36	104.2
	0.001	0.1	0.6	22.35	15.6
	0.005		1.0	9.94	14.1
	0.05		1.4	4.03	21.8

Les résultats sont basés sur 40,000 tirages Monte Carlo et 10,000 itérations Robbins-Monro. Les paramètres sont $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

Tableau 3 Réduction de Variance estimée par une méthode de Monte Carlo adptative pour un Call Européen sous le modèle de Black-Scoles

	Paramètres		**IS**		
$\boldsymbol{\gamma}$	$\boldsymbol{\sigma}$	$\mathbf{K/S_0}$	**PrixRM**	**PrixBS**	$\mathbf{RV'}$
0.005	0.3	0.6	21.59	21.60	15.3
0.05		1.0	7.12	7.12	11.1
0.1		1.2	3.45	3.45	15.0
0.5		1.6	0.67	0.67	40.7
0.004	0.1	0.6	21.45	21.46	50.6
0.5		1.0	3.40	3.40	7.8
50		1.2	0.23	0.23	31.3
500		1.4	0.004	0.004	(>1000)

Les résultats utilisent 50,000 tirages Monte Carlo. Les paramètres sont: $\mathbf{S_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

Remarque .43. Même si nous n'avons pas démontré ici des résultats théoriques qui garantissent la validité d'une telle approximation, nous pensons qu'elle devrait donner de bons résultats. Nous pensons qu'avec quelques hypothèses supplémentaires (de type "ergodicité"), l'on peut démontrer de tels résultats en adaptant (avec de légères modifications) les démonstrations des Théorèmes .13 et .14 du Chapitres 2. Dans ce dernier cas le ratio de variance que nous avons noté par $\mathbf{RV}'$ se calcule exactement de la même façon et par la même formule que **RV**. Le Tableau 3 résume les résultats numériques obtenus et confirme l'heuristique ci-dessus; ce qui montre que la méthode réduit bien la variance de l'estimation.

Chapitre 4

Efficient Variance Reduction For Functionals of Diffusions by Relative Entropy Minimization

Ce chapitre a été rédigé en collaboration avec Olivier Bardou.

4.1 Introduction

Examples of successful use of Importance Sampling applied to derivatives pricing are Reider [REI93], Glasserman, Heidelberger and Shahabuddin [GHS99a] or Boyle, Broadie and Glasserman [BBG97]. More closely related to this work are the papers by Su and Fu [SF02], Vasquez-Abad and Dufresne [VD98] and Arouna [ARO04] where stochastic approximations are used to find the optimal change of drift in an Importance Sampling procedure. In these settings, the Importance Sampling problem was based on the minimization of the variance of the estimation. In a recent work by Arouna [ARO04], such an idea was combined with the Robbins-Monro algorithm to develop an Adaptative Monte Carlo method where the optimal sampling direction is selected within the Monte Carlo iterations. The present work follows a similar idea but the main contribution is the introduction of a new minimization criterion: the Kullback-Leibler entropy (or relative entropy) between two probability measures.
Since Newton [NEW94] it is known that an Importance Sampling method can lead to a zero-variance estimate through a stochastic change of drift. And by the Girsanov Theorem, this optimal drift corresponds to an optimal probability measure. The idea is then to use the relative entropy as a distance to minimize in order to find the "best approximation" of the optimal zero-variance probability measure. We show that in a discret framework this can be achieved by using a very simple version of the Robbins-

Monro algorithm. Another nice feature of the method is that it allows to compute a consistent approximation of the delta of the derivatives.
The paper is organized as follows. We first recall the basic idea of Importance Sampling in the diffusions framework. In section 3, we show the equivalence between variance minization and entropy minimization and we investigate, as a corollary, the computation of the delta. Section 4 deals with the discrete framework for applying the method and presents the Robbins-Monro algorithm as a tool for solving the entropy minimization problem. The effiency of the method is numerically tested on a wide range of derivatives including path-dependent options and interest rate caps.

4.2 The Importance Sampling paradigm

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space and $(W_t)_{0\leq t\leq T}$ be a standard d-dimensional Brownian motion with natural filtration $\mathcal{F} := (\mathcal{F}_t)_{0\leq t\leq T}$ defined on this space. We consider the $\mathbb{R}^n$ valued process X, solution to the stochastic differential equation

$$\left\{ \begin{array}{lcl} X_0 & = & x \\ dX_t & = & b(X_t)\,dt + \sigma(X_t)dW_t,\ t \in [0,T]. \end{array} \right. \tag{.4.1}$$

We suppose that there exists two constants $K > 0$ and $C > 0$ such that the functions b and σ, defined on $\mathbb{R}^n$ and valued in $\mathbb{R}^n$ and $\mathbb{R}^{n\times d}$ respectively, verify, $\forall x,\ y \in \mathbb{R}^n$,

$$||b(x) - b(y)|| + ||\sigma(x) - \sigma(y)|| \leq K||x - y||,$$

$$||b(x)|| + ||\sigma(x)|| \leq C||x||$$

such that the SDE (.4.1) has a unique strong solution (see Karatzas and Shreve [KS88]).We are concerned with the Monte Carlo estimation of the expectation

$$V_0 := \mathbb{E}[\psi_T(X)].$$

We suppose that there exists an $\varepsilon > 0$ such that the real valued non negative functional $\psi_T(X)$ satisfies

$$\mathbb{P}\left(\psi_T(X) \geq \varepsilon\right) = 1.$$

In a financial setting, the expectation V_0 can be seen as the risk neutral price of an option with underlying asset X. The efficient computation of this price using a Monte-Carlo method often requires variance reduction techniques. The variance reduction method considered here is the so-called importance sampling method. The basic idea is as follows. Let us define

$$\mathcal{E}^T := \left\{ \theta, \, \mathcal{F}\text{-adapted such that } \mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^T \|\theta_t\|^2 \, dt \right) \right] < +\infty \right\} \tag{.4.2}$$

and

$$\mathcal{H}^T := \left\{ \theta, \, \mathcal{F}\text{-adapted such that } \mathbb{E}\left[\int_0^T \|\theta_t\|^2 \, dt \right] < +\infty \right\}. \tag{.4.3}$$

For $\theta \in \mathcal{H}^T$, the norm $\|\theta\|_T^2 := \mathbb{E}\left[\int_0^T \|\theta_t\|^2 \, dt \right]$ induces a scalar product and $\mathcal{H}^T$ endowed with this scalar product is an Hilbert space. Moreover, $\mathcal{E}^T$ is a closed non empty subset of $\mathcal{H}^T$. There is a correspondence between a process $\theta := (\theta_t)_{t \leq T} \in \mathcal{E}^T$ and $\mathbb{P}-$equivalent probability measures as stated below (see Protter [PRO90]).

Theorem 4.2.0.1 (Girsanov's Theorem)**.** *The process W^θ, defined by*

$$W_t^\theta := W_t - \int_0^t \theta_s \, ds,$$

is a standard Brownian motion under the probability $\mathbb{P}_\theta$ with density

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \exp\left(\int_0^T \theta_t \cdot dW_t - \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right)$$

with respect to $\mathbb{P}$.

The following result is a consequence of the Martingale Representation Theorem.

Theorem 4.2.0.2. *Given a probability measure $\mathbb{P}_\theta$, equivalent to $\mathbb{P}$ on $(\Omega, \, \mathcal{F}_T)$, there exists an $\mathcal{F}$-adapted process $(\theta_t)_{t \leq T}$ such that $\int_0^T \|\theta_t\|^2 dt < +\infty \quad \mathbb{P} - a.s.$ and*

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \exp\left(\int_0^T \theta_s \cdot dW_s - \frac{1}{2} \int_0^T \|\theta_s\|^2 ds \right), \quad \mathbb{P} - a.s.$$

Under the probability $\mathbb{P}_\theta$ on $(\Omega, \, \mathcal{F}_T)$, the process X is solution to the following SDE

$$\left\{ \begin{array}{rcl} X_0 & = & x \\ dX_t & = & (b(X_t) + \sigma(X_t)\theta_t) \, dt + \sigma(X_t) \, dW_t^\theta, \, t \in [0, T]. \end{array} \right. \tag{.4.4}$$

Moreover, we have

$$\mathbb{E}\left[\psi_T(X) \right] = \mathbb{E}_\theta \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right], \tag{.4.5}$$

where the subscript indicates that the expectation is computed with respect to $\mathbb{P}_\theta$. It follows that, under the probabilty $\mathbb{P}_\theta$, $\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta}$ is an unbiased estimator of the price

V_0. This is precisely an importance sampling estimate. We now have to choose the process θ that minimizes the variance of $\psi_T(X)d\mathbb{P}/d\mathbb{P}_\theta$ under $\mathbb{P}_\theta$ *i.e.* the process θ that solves

$$\min_{\theta\in\mathcal{E}^T} \mathbb{V}\mathrm{ar}_\theta\left[\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right]. \tag{.4.6}$$

Theoretically, it is known that there exists a unique probability measure $\mathbb{P}_{\theta^*}$ for which $\mathbb{V}\mathrm{ar}_{\theta^*}\left[\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_{\theta^*}}\right]$ vanishes. In fact, since

$$\begin{aligned}\mathbb{V}\mathrm{ar}_\theta\left[\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right] &= \mathbb{E}_\theta\left[\left(\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right)^2\right] - \mathbb{E}_\theta\left[\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right]^2\\ &= \mathbb{E}\left[\psi_T(X)^2\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right] - \mathbb{E}\left[\psi_T(X)\right]^2,\end{aligned}$$

$\mathbb{P}_{\theta^*}$ should satisfy

$$\mathbb{E}\left[\psi_T(X)^2\frac{d\mathbb{P}}{d\mathbb{P}_{\theta^*}}\right] = \mathbb{E}\left[\psi_T(X)\right]^2,$$

and this relation leads obviously to

$$\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}} = \frac{\psi_T(X)}{\mathbb{E}\left[\psi_T(X)\right]},$$

which is equivalent to

$$d\mathbb{P}_{\theta^*} = \frac{\psi_T(X)}{V_0}d\mathbb{P}. \tag{.4.7}$$

Of course, this formula is of no direct practical interest since it requires the knowledge of the expectation V_0. Our aim is now to built an efficient estimator of the process θ^*. This problem will be formulated as a distance minimization issue. We are going to see that a convenient choice of the distance is important from a numerical point of vue. The next section is thus devoted to the comparison between different distances.

4.3 Variance reduction as entropy minimization

4.3.1 χ^2 minimization

The χ^2 divergence between two probability measures is formally defined by

$$\begin{aligned}\mathcal{R}_2\left(\mathbb{P},\,\mathbb{Q}\right) &= \frac{1}{2}\int\left(\frac{d\mathbb{P}}{d\mathbb{Q}}-1\right)^2 d\mathbb{Q} \text{ if } \mathbb{P}\sim\mathbb{Q}\\ &= +\infty \text{ otherwise.}\end{aligned}$$

When considering this distance, simple computations show that problem (.4.6) is equivalent to the following distance minimization problem between $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta^*}$

$$\min_{\theta \in \mathcal{E}^T} \mathcal{R}_2 \left(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta \right). \tag{.4.8}$$

To see this, first observe that

$$\begin{aligned} \mathcal{R}_2 \left(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta \right) &= \frac{1}{2} \int \left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_\theta} - 1 \right)^2 d\mathbb{P}_\theta \\ &= \frac{1}{2} \left(\int \frac{\left(d\mathbb{P}_{\theta^*} \right)^2}{d\mathbb{P}_\theta} - 1 \right). \end{aligned}$$

Plugging equation (.4.7) in this expression gives

$$\begin{aligned} \mathcal{R}_2 \left(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta \right) &= \frac{1}{2} \left(\int \frac{\psi_T(X)^2}{\mathbb{E}\left[\psi_T(X) \right]^2} \frac{d\mathbb{P}}{d\mathbb{P}_\theta} d\mathbb{P} - 1 \right) \\ &= \frac{1}{2} \left(\frac{\mathbb{E}\left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right]}{V_0^2} - 1 \right). \end{aligned}$$

Therefore, problem (.4.8) reexpresses as

$$\min_{\theta \in \mathcal{E}^T} \mathbb{E}\left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right]. \tag{.4.9}$$

By the Girsanov Theorem this latter problem reduces to

$$\min_{\theta \in \mathcal{E}^T} \mathbb{E}\left[\psi_T(X)^2 \exp\left(- \int_0^T \theta_t \cdot dW_t + \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right) \right]. \tag{.4.10}$$

Equation (.4.9) is exactly the usual criterion to minimize when refering to optimal importance sampling. See for instance Newton [NEW94] and Arouna [ARO04]. From this latter reference, we know that considering the χ^2 distance leads to numerical difficulties. In particular, it is shown that a truncated Robbins-Monro algorithm has to be introduced in order to deal with the exponential martingale. We are now going to see how a another choice of distance can overcome this difficulty.

4.3.2 Kullback-Leibler minimization

This distance is defined by

$$\begin{aligned} \mathcal{R}_0 \left(\mathbb{P}, \mathbb{Q} \right) &= \int \ln \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{P} \text{ if } \mathbb{P} \sim \mathbb{Q} \\ &= +\infty \text{ otherwise.} \end{aligned}$$

Our aim is to minimize $\mathcal{R}_0\left(\mathbb{P}_{\theta^*},\, \mathbb{P}_\theta\right)$. This quantity can be rewritten as

$$\begin{aligned}\mathcal{R}_0\left(\mathbb{P}_{\theta^*},\, \mathbb{P}_\theta\right) &= \int \ln\left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_\theta}\right) d\mathbb{P}_{\theta^*} \\ &= \int \ln\left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}} \cdot \frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right) d\mathbb{P}_{\theta^*} \\ &= \int \ln\left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}}\right) d\mathbb{P}_{\theta^*} + \int \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right) d\mathbb{P}_{\theta^*}.\end{aligned}$$

The first integral does not depend on $\mathbb{P}_\theta$. Then, minimizing $\mathcal{R}_0\left(\mathbb{P}_{\theta^*},\, \mathbb{P}_\theta\right)$ is equivalent to minimizing $\int \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right) d\mathbb{P}_{\theta^*}$. Moreover, since

$$d\mathbb{P}_{\theta^*} = \frac{\psi_T(X)}{\mathbb{E}\left[\psi_T(X)\right]} d\mathbb{P},$$

we get,

$$\begin{aligned}\int \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right) d\mathbb{P}_{\theta^*} &= \int \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right) \frac{\psi_T(X)}{\mathbb{E}\left[\psi_T(X)\right]} d\mathbb{P} \\ &= \frac{\mathbb{E}\left[\psi_T(X) \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right)\right]}{\mathbb{E}\left[\psi_T(X)\right]}.\end{aligned}$$

Thus, the problem of minimizing $\mathcal{R}_0\left(\mathbb{P}_{\theta^*},\, \mathbb{P}_\theta\right)$ reduces to

$$\min_{\theta \in \mathcal{E}^T} \mathbb{E}\left[\psi_T(X) \ln\left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta}\right)\right], \tag{.4.11}$$

which we rewrite as, by the virtue of the Girsanov Theorem,

$$\min_{\theta \in \mathcal{E}^T} J(\theta), \tag{.4.12}$$

with

$$J(\theta) := \mathbb{E}\left[\psi_T(X)\left(-\int_0^T \theta_t \cdot dW_t + \frac{1}{2}\int_0^T \|\theta_t\|^2 dt\right)\right]. \tag{.4.13}$$

The following result is important for the understanding of the numerical implementation of our method.

Theorem 4.3.2.1. *Suppose that K^T is a closed, convex and non empty subset of $\mathcal{E}^T$. If there exists $\varepsilon > 0$ such that $\mathbb{P}[\psi_T(X) \geq \varepsilon] = 1$ then J admits a unique minimum $\theta^* \in K^T$ and*

$$\|\theta - \theta^*\|_T^2 \leq \frac{4}{\varepsilon}\left[J(\theta) - J(\theta^*)\right], \quad \forall\, \theta \in K^T. \tag{.4.14}$$

Remark 4.3.1. For a proof, one could refer to any book of convex analysis.

4.3.3 Comparison between these distances

First, notice that the pseudo-distances $\mathcal{R}_2(\mathbb{P}, \mathbb{Q})$ and $\mathcal{R}_0(\mathbb{P}, \mathbb{Q})$ are non negative and vanish if and only if $\mathbb{P} = \mathbb{Q}$. Therefore, they are equivalent in terms of zero-minimization. However, the Kullback Leibler distance is sharper than the χ^2 divergence in the sense that

$$\mathcal{R}_0(\mathbb{P}, \mathbb{Q}) \leq \mathcal{R}_2(\mathbb{P}, \mathbb{Q}), \ \forall \mathbb{P} \sim \mathbb{Q}, \tag{.4.15}$$

since $\ln x \leq \frac{1}{2}(x^2 - 1)$, $\forall x > 0$. But, the most important fact is the quadratic dependance of the right hand side of equation (.4.12) instead of the exponential dependance of equation (.4.10) upon θ. Clearly, (.4.12) leads to a more tractable minimization procedure than (.4.10). We refer to Basseville (1996) for complements on entropy and pseudo-distances between probability measures.

4.3.4 Delta computation

As noted in the introduction, a key feature of importance sampling for pricing a given derivative is that the optimal control θ^* is closely related to the delta of this derivative. In fact, it is known from Newton [NEW94] that the zero-variance drift term θ^* of an option with discounted payoff ψ_T is given by

$$\theta_t^* = \frac{\sigma(X_t)\pi_t}{\mathbb{E}[\psi_T(X)|\mathcal{F}_t]}, \ t \in [0, T]$$

where the process π_t is the *unique* process implicitely defined by the Martingale Representation theorem

$$\psi_T(X) = \mathbb{E}[\psi_T(X)] + \int_0^T \pi_t \cdot \sigma(X_t)\, dW_t, \quad a.s. \tag{.4.16}$$

The process π is refered to as the delta of the option with payoff $\psi_T(X)$. Thus, the computation of the optimal process θ^* gives also an estimation of the delta of the option through the relation :

$$\pi_t = \mathbb{E}[\psi_T(X)|\mathcal{F}_t]\, \sigma^{-1}(X_t) \cdot \theta_t^* \quad a.s. \tag{.4.17}$$

which reduces to

$$\Delta := V_0 \sigma^{-1}(x) \cdot \theta_0^*, \tag{.4.18}$$

for $t = 0$. We now develop the discrete framework to perform the computation of θ^* and V_0.

4.4 The Variance Reduction Procedure

In practice, when the solution to the SDE (.4.1) can not be computed explicitly, one must discretize it. In such a case the computation of the price V_0 reduces to the computation of

$$\hat{V_0} = \mathbb{E}\left[\psi_T(\bar{X}^h)\right],$$

where $\bar{X}^h$ is a discretization of the solution X to (.4.1). For sake of simplicity, we shall focus on the Euler-Maruyama scheme (see Talay and Tubaro [TAL90]) although our method could be extended to any other scheme. The discretization of equation (.4.1) with constant time step $h := \frac{T}{m}$ is then

$$\left\{ \begin{array}{lcl} \bar{X}_0 & = & X_0 \\ \bar{X}_{h(p+1)} & = & \bar{X}_{hp} + b(\bar{X}_{hp})h + \sigma(\bar{X}_{hp})\sqrt{h}\Delta W_p,\ p = 0 \ldots m-1 \end{array} \right.$$

with $\Delta W_p \overset{iid}{\sim} \mathcal{N}(0, I_d)$ and I_d the identity matrix of $\mathbb{R}^d$. Similarly, equation (.4.4) is discretized by the following scheme :

$$\left\{ \begin{array}{lcl} \tilde{X}_0 & = & X_0 \\ \tilde{X}_{h(p+1)} & = & \tilde{X}_{hp} + \left(b(\tilde{X}_{hp}) + \sigma\left(\tilde{X}_{hp}\right)\tilde{\theta}_p\right)h + \sigma(\tilde{X}_{hp})\sqrt{h}\Delta W_p,\ p = 0 \ldots m-1. \end{array} \right.$$

In this representation, $\tilde{\theta} := \left(\tilde{\theta}_0, \ldots, \tilde{\theta}_{m-1}\right)$ is a $d \times m$ matrix. Let us define the picewise constant function $(\tilde{\theta}_t)_{t \leq T}$ by setting

$$\tilde{\theta}_t := \tilde{\theta}_p, \quad \text{if} \quad t \in [hp,\ h(p+1)[.$$

Then $\tilde{\theta}$ can be seen as a rough approximation of the random vector process θ. In the sequel ΔW denotes the $d \times m$ matrix $(\Delta W_0, \ldots, \Delta W_{m-1})$. In this discrete setting, equation (.4.5) translates into

$$\begin{aligned} \hat{V_0} &= \mathbb{E}\left[\psi_T(\bar{X}\right] \\ &= \mathbb{E}\left[\psi_T(\tilde{X})\exp\left(-\tilde{\theta}\cdot\sqrt{h}\Delta W - \frac{h}{2}\|\tilde{\theta}\|^2\right)\right] \end{aligned} \tag{.4.19}$$

where $\|M\| = \left(\sum_{p=1}^{m}\sum_{k=1}^{d} M_{kp}^2\right)^{1/2}$ denotes the Frobenius norm of a matrix M of size $d \times m$. These relations hold for each $\tilde{\theta} \in \mathbb{R}^{d\times m}$. From equation (.4.12), the Kullback-Leibler minimization becomes

$$\min_{\tilde{\theta}\in\mathcal{E}^T} \tilde{J}(\tilde{\theta}), \quad \tilde{\theta} \in \mathbb{R}^{d\times m}, \tag{.4.20}$$

with

$$\tilde{J}(\tilde{\theta}) := \mathbb{E}\left[\psi_T(\bar{X})\left(-\tilde{\theta}\cdot\sqrt{h}\Delta W + \frac{h}{2}\|\tilde{\theta}\|^2\right)\right].$$

Using first order optimality conditions, we are reduced to solving the following zero-problem

$$\nabla\tilde{J}(\tilde{\theta}) = 0, \quad \tilde{\theta}\in\mathbb{R}^{d\times m}, \tag{.4.21}$$

with

$$\nabla\tilde{J}(\tilde{\theta}) := \mathbb{E}\left[\psi_T(\bar{X})\left(h\tilde{\theta} - \sqrt{h}\Delta W\right)\right].$$

Remark 4.4.1. The functional $\tilde{J}(\tilde{\theta})$ is defined as an expectation under the initial probability so that the functional $\psi_T(\bar{X})$ does not depend upon $\tilde{\theta}$ which only appears through the functional $(-\tilde{\theta}\cdot\sqrt{h}\Delta W + \frac{h}{2}\|\tilde{\theta}\|^2)$ which is clearly convex. As this functional explodes when $\|\tilde{\theta}\|\to+\infty$, it admits a unique minimum $\tilde{\theta}^* \in \mathbb{R}^{d\times m}$.
As the exponential is a convex application, the same arguments can be used to prove that the χ^2 distance is also a convex functional of $\tilde{\theta}$ (see Arouna [ARO04]). Moreover, remember that the χ^2 distance is always greater than the Kullback-Leibler distance (see (.4.15)). This feature is illustrated in figure 4.1. From this, it is easy to check that an optimization algorithm based on the Kullback-Leibler distance always leads to a decrease in the variance, even if the optimal process θ^* is not reached.

4.4.1 The Martingale Monte Carlo Algorithm

Following Arouna [ARO04], the Monte Carlo method to be implemented in this case can be described as follows. First suppose that we are able to compute efficiently a sequence $(\tilde{\theta}^i)_{i\geq 0}$ of approximations of the optimal $\tilde{\theta}^*$ generated from a sequence of independent copies $(\Delta W^i)_{i\geq 1}$ of ΔW. Then using (.4.19) and a result of Arouna [ARO04] we have

$$\tilde{V}_N := \frac{1}{N}\sum_{i=1}^{N}\psi_T(\tilde{X}^i)\exp\left(-\tilde{\theta}^i\cdot\sqrt{h}\Delta W^i - \frac{h}{2}\|\tilde{\theta}^i\|^2\right) \underset{N\to+\infty}{\overset{a.s.}{\longrightarrow}} \mathbb{E}\left[\psi_T(\bar{X})\right], \tag{.4.22}$$

where $(\tilde{X}^i)_{i=1\ldots N}$ are independent copies of $\tilde{X}$ simulated with ΔW^i and $\tilde{\theta}^i$. Note that no restriction is assumed on the dependence between this sequence $\tilde{\theta}^i$ and the random process ΔW^i. The variance of the estimation is given by

$$\Sigma_N^2 := \frac{1}{N}\sum_{i=1}^{N}\psi_T^2(\tilde{X}^i)\exp\left(-2\tilde{\theta}^i\cdot\sqrt{h}\Delta W^i - h\|\tilde{\theta}^i\|^2\right) - \tilde{V}_N^2, \tag{.4.23}$$

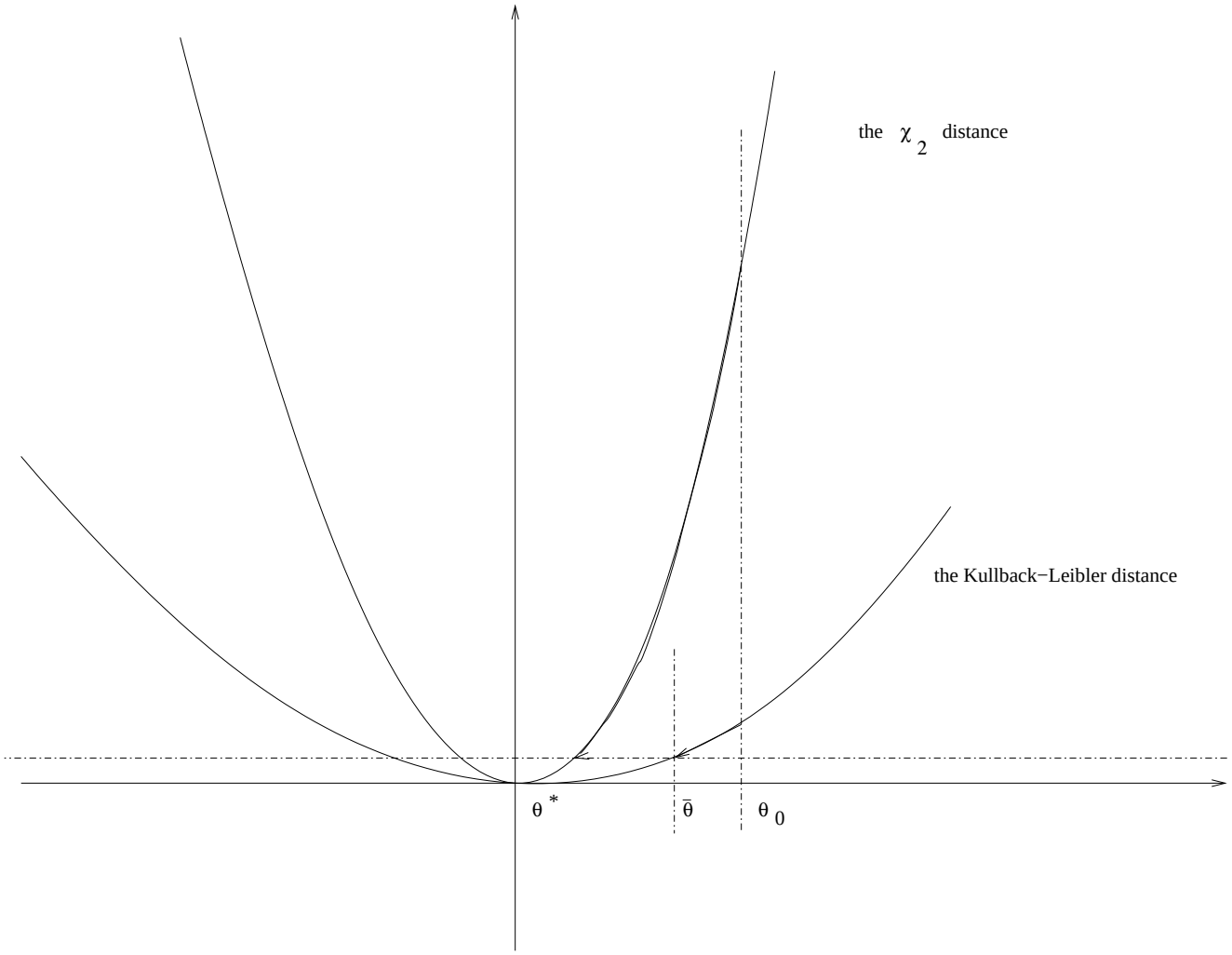


Figure 4.1: Variance and cross-entropy minimization

and it is proved in Arouna [ARO04] that, on one hand,

$$\Sigma_N^2 \underset{N\to+\infty}{\overset{a.s.}{\longrightarrow}} \mathbb{V}\mathrm{ar}_{\tilde{\theta}^*}\left[\psi_T(X)\frac{d\mathbb{P}}{d\mathbb{P}_{\tilde{\theta}^*}}\right] \tag{.4.24}$$

and, on the other hand,

$$\sqrt{N}\frac{\tilde{V}_N - \mathbb{E}\left[\psi_T(\bar{X})\right]}{\Sigma_N^2} \underset{N\to+\infty}{\overset{\text{Law}}{\longrightarrow}} \mathcal{N}(0,\,1)\,. \tag{.4.25}$$

Thus, we have a central limit theorem that enables us to build confidence intervals. Since the asymptotic variance of the estimation (.4.22) convergences to a variance smaller than that of a crude Monte-Carlo simulation, it can be shown that for N large enough Σ_N^2 is also smaller than the variance of a crude Monte-Carlo simulation.
In addition, we can get as a feedback a good approximation to the delta of the option via formula (.4.18) by setting

$$\hat{\Delta} := \tilde{V}_N \sigma^{-1}(0,\,x)\cdot\tilde{\theta}_0^N\,. \tag{.4.26}$$

Next the minimization issue (.4.21) will be performed using a simple Robbins-Monro procedure as described below.

4.4.2 A stochastic approximation scheme to learn the optimal drift

From the discretization of X, we see that the simulation of $\bar{X}$ relies only upon the simulation of the law of $\Delta W \sim \mathcal{N}(0,\,I_{d\times m})$. Hence there exists a function $\varphi_T(x)$ depending on the discretization of X such that

$$\psi_T(\bar{X}) = \varphi_T(\Delta W).$$

Let us define

$$F(\theta,\,x) := \varphi_T(x)(\sqrt{h}\theta - x)$$

and suppose that

$$(H_0):\ \exists\,\alpha>0 \quad \text{such that} \quad \mathbb{E}\left[\varphi_T^{2+\alpha}(\Delta W)\right] < +\infty.$$

Our approach for solving problem (.4.21) relies on the following iterative scheme

$$\tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma_{i+1}F(\tilde{\theta}^i,\,\Delta W^i), \tag{.4.27}$$

where $(\gamma_i)_{i\geq 1}$ is a sequence of positive numbers converging to 0. This iterative stochastic approximation in known as Robbins–Monro algorithm and is well indicated for the resolution of the zero-search problem

$$\mathbb{E}[F(\tilde{\theta}^*, \Delta W)] = 0, \quad \tilde{\theta}^* \in \mathbb{R}^{d\times m}. \tag{.4.28}$$

This algorithm is particularly efficient when the expectation $\mathbb{E}[F(\tilde{\theta}, \Delta W)]$ cannot be computed but one can easily simulate the law of $F(\tilde{\theta}, \Delta W)$. Let us now denote by $\left(\tilde{\mathcal{F}}_i\right)_{i=1\ldots n}$ the σ-algebra generated by the random vectors $(\Delta W^i)_{i=1\ldots n}$. If we set

$$E(\tilde{\theta}) := \mathbb{E}[F(\tilde{\theta}, \Delta W)],$$

clearly we have

$$\mathbb{E}[F(\tilde{\theta}^i, \Delta W^i)|\tilde{\mathcal{F}}_{i-1}] = E(\tilde{\theta}^i),$$

and

$$\mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2|\tilde{\mathcal{F}}_{i-1}] \leq C(1 + \|\tilde{\theta}^i\|^2), \quad a.s. \tag{.4.29}$$

where

$$C = \max\left(\mathbb{E}[\varphi_T^2(\Delta W)], \mathbb{E}[\varphi_T^2(\Delta W)\|\Delta W\|^2]\right)$$

is a finite positive constant by the virtue of hypothesis (H_0). Next, we state a convergence Theorem for (.4.27) proved in Duflo [DUF97]:

Theorem 4.4.2.1. *Under the following hypotheses*

$$\exists!\tilde{\theta}^* \in \mathbb{R}^{d\times m} \text{ such that } E(\tilde{\theta}^*) = 0 \text{ and } \forall\tilde{\theta} \in \mathbb{R}^{d\times m} \text{ with } \tilde{\theta} \neq \tilde{\theta}^*, (\tilde{\theta} - \tilde{\theta}^*)\cdot E(\tilde{\theta}) > 0, \tag{.4.30}$$

$$\sum_i \gamma_i = +\infty \quad \text{and} \quad \sum_i \gamma_i^2 < +\infty, \tag{.4.31}$$

$$\mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2|\tilde{\mathcal{F}}_{i-1}] < C(1 + \|\tilde{\theta}^i\|^2), \, a.s., \tag{.4.32}$$

the sequence of random vectors $(\tilde{\theta}^i)_{i\geq 0}$ defined in (.4.27) converges almost surely and in L^2 to $\tilde{\theta}^*$.

Since $\tilde{J}$ is convex, it is clear that there exists a unique $\tilde{\theta}^*$ which minimizes the functional $\tilde{J}(\tilde{\theta})$ so that hypothesis (.4.30) is a simple consequence of the Theorem of finite increments. Since hypothesis (.4.32) is fulfilled from equation (.4.29), Theorem (4.4.2.1) applies for the iterative algorithm (.4.27).
This algorithm is adaptive in the sense that the sequence $\tilde{\theta}$ is built up within the Monte-Carlo algorithm. As this sequence is convergent it admits implicit bounds. Nevertheless, numerical events might lead to large values of $\tilde{\theta}$ and make our exponential terms unstable. We shall adress this issue in the next section.

4.4.3 Summary

Finally, let us summarize the Martingale Monte Carlo algorithm to be implemented in the very general case.

$$\tilde{\theta}^1 = \theta^0,\ h = \frac{T}{p},\ \gamma^i \xrightarrow[i\to+\infty]{} 0,\ \sum_{i=1}^{+\infty} \gamma^i = +\infty.$$

for $i = 1 \dots m$,

$$\bar{X}_0^i = X_0,\ \tilde{X}_0^i = X_0,\ \Delta W^i \sim \mathcal{N}\left(0,\ I_{d,p}\right)$$

for $j = 1 \dots p$,

$$\bar{X}_j^i = \bar{X}_{j-1}^i + b(\bar{X}_{j-1}^i)h + \sigma(\bar{X}_{j-1}^i) \cdot \sqrt{h}\Delta W_j^i$$

$$\tilde{X}_j^i = \tilde{X}_{j-1}^i + \left(b(\tilde{X}_{j-1}^i) + \sigma(\tilde{X}_{j-1}^i) \cdot \tilde{\theta}_{j-1}^i\right) h + \sigma(\tilde{X}_{j-1}^i) \cdot \sqrt{h}\Delta W_j^i$$

$$\tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma^{i+1}\psi_T\left(\bar{X}^i\right)\left(\sqrt{h}\tilde{\theta}^i - \Delta W^i\right)$$

$$\tilde{V}_N := \frac{1}{N}\sum_{i=1}^{N} \psi_T\left(\tilde{X}^i\right) \exp\left(-\sqrt{h}\,\mathrm{Tr}\left(\tilde{\theta}^i \cdot {}^t\Delta W^i\right) - \frac{h}{2}||\tilde{\theta}^i||_F^2\right)$$

Remark 4.4.2. Of course, the Robbins-Monro algorithm and the Monte-Carlo procedure can also be performed separately.

4.5 Numerical Results

4.5.1 Preliminary remarks

Choice of the sequence γ_i

It is worth pointing out that the Robbins–Monro algorithm (.4.27) is very easy to implement. In our numerical investigations, we use $\gamma_i = \alpha/(1+i)$ as sequence of convergent steps, where $\alpha > 0$ is chosen empirically. In general this choice is not an easy task. Hence we followed the numerical recommandations for financial applications that are made at section 6 of Arouna [ARO04]. For a more global insight into this problem, one can see Kushner and Yin [KY03], where the optimal numerical convergence of the Robbins–Monro algorithm is discussed in length.

Reduction of the conditional variance

Remember that the observation of the the Robbins-Monro algorithm (.4.27) is defined as

$$F(\tilde{\theta}^i, \Delta W^i) = \varphi_T(\Delta W^i)(\sqrt{h}\tilde{\theta}^i - \Delta W^i), \quad i \geq 0.$$

In order to reduce the variability of the Robbins-Monro algorithm, one might reduce the conditional variance $\mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2 | \tilde{\mathcal{F}}_i]$ of the sequence of the observations. This can be done by replacing in equation (.4.27) the observation $F(\tilde{\theta}^i, \Delta W^i)$ with $\hat{F}(\tilde{\theta}^i, \Delta W^i)$ where

$$\hat{F}(\tilde{\theta}, \Delta W) = \frac{1}{2}\left(F(\tilde{\theta}, \Delta W) + F(\tilde{\theta}, -\Delta W)\right).$$

Clearly we still have

$$\mathbb{E}[\hat{F}(\tilde{\theta}, \Delta W)] = 0,$$

and Theorem (4.4.2.1) still applies for the Robbins-Monro algorithm defined as

$$\tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma_{i+1}\hat{F}(\tilde{\theta}^i, \Delta W^i), \tag{.4.33}$$

Control of the exponential terms

Since this algorithm converges, it admits implicit bounds. But from a numerical point of view, it is still possible for the $\tilde{\theta}^i$ to take excessive values. This might be a consequence of the γ_i that are too large or of numerical events. To deal with these problems, we implement the following truncated version of the algorithm

$$\tilde{\theta}^{i+1} = \begin{cases} \tilde{\theta}^i - \gamma_{i+1}\hat{F}(\tilde{\theta}^i, \Delta W^i) & \text{if} \quad \left\|\tilde{\theta}^i - \gamma_{i+1}\hat{F}(\tilde{\theta}^i, \Delta W^i)\right\| \leq U, \\ \tilde{\theta}^i & \text{otherwise.} \end{cases} \tag{.4.34}$$

U is a constant upperbound of $\tilde{\theta}^*$ taken equal to 5 in our all experimentations.

4.5.2 European Call

We begin our experiments with a european Call option example in the Black–Scholes model. Thus, our discretized process follows the dynamic

$$\bar{X}_{t_i} = \bar{X}_{t_{i-1}} e^{(r-\sigma^2/2+\tilde{\theta}_{t_{i-1}})(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}}\Delta W_i}, \; \bar{X}_0 = 1, \; t_i = \frac{i}{n}, \; i = 1 \ldots n. \tag{.4.35}$$

The pricing and the hedging of this option is available in closed form but this example is still illustrative by allowing us to measure both the variance reduction obtained and

the accuracy in prices and in delta hedging. The variance reduction is measured via the ratio

$$\mathbf{RV} := \frac{\frac{1}{N}\sum_{i=1}^{N}\psi_T^2(\bar{X}^i) - \left(\frac{1}{N}\sum_{n=1}^{N}\psi_T(\bar{X}^i)\right)^2}{\Sigma_N^2},$$

where Σ_i^2 is defined in (.4.23). The numerical results for this case are in Table 1 and show the effectiveness of the procedure. **PrixRM**, **PrixBS**, $\mathbf{\Delta_{RM}}$ and $\mathbf{\Delta_{BS}}$ denote respectively the Monte Carlo estimated price including our method (Importance Sampling + Relative entropy minimization + Robbins–Monro), the Black–Scholes exact price, the delta computed by our method (using equation (.4.26)) and the exact delta given by the Black–Scholes formula.

Table 1 Estimated variance reduction ratio and delta hedging for a european call option in the Black–Scholes model

Parameters						RM+IS	
$\boldsymbol{\alpha}$	$\boldsymbol{\sigma}$	$\mathbf{K/S_0}$	**PrixRM**	$\mathbf{\Delta_{RM}}$	**PrixBS**	$\mathbf{\Delta_{BS}}$	**RV**
1	0.1	0.6	21.46	1.00	21.46	1.00	104.1
20		1.0	3.41	0.71	3.40	0.71	7.7
5000		1.4	0.004	0.002	0.004	0.002	748.5
2.5	0.3	0.6	21.61	0.97	21.60	0.99	16.5
25		1.0	7.13	0.62	7.12	0.62	11.3
50		1.4	1.56	0.21	1.56	0.21	25.4

The results are based on 40,000 Monte Carlo simulated paths.
The parameters are: $\mathbf{\bar{X}_0 = 50},\ \mathbf{r = 0.05},\ \mathbf{T = 1.0}$.

4.5.3 Asian Call

Our second example is the pricing of an arithmetic asian call option in the Black–Scholes model. The discounted payoff of this option is given by

$$\varphi(\bar{X}) := e^{-rT}\max\left(\frac{1}{n}\sum_{i=1}^{n}\bar{X}_{t_i} - K,\, 0\right)$$

where n is the total number of reset dates and $\bar{X}$ is defined in (.4.35). Numerical procedures are available for the pricing of this option with continuous averaging but not with discret averaging (see Geman and Yor [GEM93]). In this latter case, simulation is

necessary. The numerical results obtained are summarized in Table 2. $\mathbf{\Delta_{DF}}$ represents the delta of the option computed with a finite differences method (with step $h = 0.01$). Note the strong accuracy of the delta portofolio computed with our algorithm.

Table 2 Estimated variance reduction ratio and delta hedging for an arithmetic asian call option in the Black–Scholes model

Parameters					RM+IS		
n	$\boldsymbol{\alpha}$	$\boldsymbol{\sigma}$	$\mathbf{K/S_0}$	**PrixRM**	$\mathbf{\Delta_{RM}}$	$\mathbf{\Delta_{BS}}$	**RV**
16	5	0.3	0.8	11.09	0.91	0.91	9.8
	10		1.0	4.18	0.59	0.57	9.4
	50		1.2	1.09	0.21	0.22	18.6
16	10	0.1	0.8	10.80	0.97	0.98	36.7
	25		1.0	1.92	0.67	0.67	7.2
	50		1.1	0.20	0.13	0.13	12.3
64	5	0.3	0.8	10.93	0.92	0.92	9.2
	20		1.0	3.98	0.58	0.58	9.0
	50		1.2	0.97	0.22	0.22	18.4
64	5	0.1	0.8	10.73	0.99	0.98	27.6
	25		1.0	1.83	0.67	0.66	6.8
	500		1.1	0.17	0.13	0.13	21.1

The results are based on 10,000 Monte Carlo simulated paths.
The parameters are: $\mathbf{\bar{X}_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

4.5.4 European Basket Call

Our last example deals with the pricing of a european basket call option. The payoff of this option is given by

$$\max\left(\sum_{i=1}^{d} a_i \bar{X}_T^i - K,\, 0\right),$$

where a_i are positive constants such that $\sum_{i=1}^{d} a_i = 1$. The dynamic of the i^{th} asset $\bar{X}^i$ (under the so-called risk neutral probability) is defined as

$$\frac{d\bar{X}_t^i}{\bar{X}_t^i} = (r + \sum_{j=1}^{d} \sigma_{ij}\tilde{\theta}(t)_j^i)dt + \sum_{j=1}^{d} \sigma_{ij} dW_t^j, \quad \bar{X}_0^i > 0,$$

with $W = (W^1, \ldots, W^d)$ a $d-$dimensional Brownian motion, $\tilde{\theta}(t)^i_j$ a deterministic $d \times d$-sized matrix representing the Girsanov change of drift and $\sigma = (\sigma_{ij})_{ij}$ a $d \times d$-sized matrix representing the volatility of the market. Once again, simulation is the only method available for pricing, especially when d is large ($d \geq 10$). For simplicity, we fix the spot values of all the underlying assets equal to $\bar{X}_0$. The volatility matrix is flat at 10 % or 30 %. Table 3 displays the variance reduction obtained. Observe that in the worst case we have $\sqrt{\mathbf{RV}} > 3$. This means that we are able to reduce the length of the Monte Carlo confidence intervals (at a given confidence level) by a factor of at least 3. As mentioned above, our method also provide an estimate of the delta hedging strategy. Under flat spot values and volatility matrix hypothesis, it can be proved that the pricing problem reduces to the pricing of a simple european call option on an underlying $\bar{X}$ that follows the dynamic

$$\frac{d\bar{X}_t}{\bar{X}_t} = (r + \sigma\tilde{\theta}(t))dt + \sigma\sqrt{d}\, dB_t, \quad \bar{X}_0 > 0,$$

with $\sigma = \sigma_{ij}, \quad \forall\, i,\, j$ and $B_t := \frac{1}{\sqrt{d}} \sum_{j=1}^{d} W_t^j$, a one dimensional Brownian motion. Thus the delta hedging portofolio of the option is given by the so-called Black–Scholes formula.
For example if $n = 10$ assets, $\bar{X}_0 = K = 50$, $\sigma = 30\,\%$, then the seller of the option must hold $\mathbf{\Delta_{BS}} = 0.69$ share of each underlying asset. With these values as inputs, we obtained the following hedging portofolio

$$\mathbf{\Delta_{RM}} = \{0.73,\, 0.69,\, 0.67,\, 0.69,\, 0.70,\, 0.71,\, 0.70,\, 0.72,\, 0.70,\, 0.72\}\,,$$

after $RM = 20,000$ iterations of the Robbins-Monro algorithm (.4.34). This shows the potential of the procedure even in high dimensional cases.

Table 3 Estimated variance reduction ratio for a european basket call option in the Black–Scholes model

	Parameters			RM+IS	
n	$\boldsymbol{\alpha}$	$\boldsymbol{\sigma}$	$\mathbf{K/\bar{X}_0}$	**PrixRM**	**RV**
10	2.5	0.3	0.6	27.22	34.8
	5		1.0	19.01	34.2
	10		1.4	13.83	38.7
	2.5	0.1	0.6	21.62	14.8
	10		1.0	7.41	11.2
	25		1.4	1.79	23.3
20	1.5	0.3	0.6	31.74	99.5
	2.5		1.0	25.59	94.5
	2.5		1.4	21.35	99.4
	2.5	0.1	0.6	22.34	15.1
	2.5		1.0	9.94	13.9
	5		1.4	4.03	21.0

The results are based on 40,000 Monte Carlo simulated paths.
The parameters are: $\mathbf{\bar{X}_0 = 50}$, $\mathbf{r = 0.05}$, $\mathbf{T = 1.0}$.

In all these examples the additional computing time per replication required to simulate using Robbins-Monro algorithm is roughly (at most) about 30 % of the time required to simulate without variance reduction. But this additional effort is negligible when compared with the variance reduction achieved and as a feedback the method provides accurate estimate of the delta of the option.

4.6 Conclusion

We present in this paper a variance reduction method based on the relative entropy minimization between an optimal Importance Sampling measure and a set of parametric Importance Sampling measures. The optimal Importance Sampling measure corresponds to an optimal drift parameter which itself provides the optimal hedging strategy for the option (see Newton[NEW94]). We use the Robbins-Monro algorithms to find a good asymptotic estimate of this optimal drift and thus for the delta of the option. Our numerical tests indicate the effectiveness of the procedure, since the variance reductions achieved are important. The method is as easy to implement as is a standard Monte Carlo method. One can take advantage of this facility to combine the

method with other variance reduction techniques for more efficiency.

Bibliographie

[AJ99] AVEN T. and JENSEN U. *Stochastic Models in Reliabilty*. Springer, 1999.

[ARO04] AROUNA B. Robbins-monro algorithms and variance reduction in finance. *The Journal of Computational Finance*, 7(2), 2003/04.

[ARO04] AROUNA B. Adaptative monte carlo method, a variance reduction technique. *Monte Carlo Methods And Applications*, 10(1), 2004.

[BBG97] BOYLE P., BROADIE M., and GLASSERMAN P. Monte carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1257–1321, 1997.

[BLL86] BOULEAU N., LAMBERTON D., and LAPEYRE B. A numerical method suited to the probabilistic design of structure. *Journal of theoretical and applied mechanics*, 5:781–801, 1986.

[BMP87] BENVENISTE A., METIVIER M., and PRIOURET P. *Algorithmes adaptatifs et approximations stochastiques*. Masson, 1987.

[CGG88] CHEN H-F., GUO L., and GAO A-J. Convergence and robustness of the robbins-monro algorithm truncated at randomly varying bounds. *Stochastic Processes and their Applications*, 27:217–231, 1988.

[CHE02] CHEN H-F. *Stochastic Approximation and Its Applications*. Kluwer Academic Publisher, 2002.

[CLA84] CLARK D.S. Necessary and sufficient conditions for the robbins-monro method. *Stochastic Processes and their Applications*, 17:359–367, 1984.

[CZ86] CHEN H-F. and ZHU Y.M. Stochastic approximation procedure with randomly varying truncations. *Scientia Sinica (series A)*, 29:914–926, 1986.

[DD83] DACUNHA-CASTELLE D. and DUFLO M. *Probabilités et Statisques (tome 2.)*. Masson, France, 1983.

[DEL96] DELYON B. General results on the convergence of stochastic algorithms. *IEEE-A.C*, 41(9):1245–1255, 1996.

[DEL00] DELYON B. *Stochastic approximation with decreasing gain: Convergence and asymptotic theory.* DESS course, INRIA-IRISA, 2000.

[DUF96] DUFLO M. *Algorithmes stochastiques.* Springer-Verlag, France, 1996.

[DUF97] DUFLO M. *Random Iterative Models.* Springer-Verlag, France, 1997.

[GEM93] GEMAN H. and YOR M. Bessel processes, asian options, and perpetuities. *Mathematical Finance*, 3, 1993.

[GHS99a] GLASSERMAN P., HEIDELBERGER P., and SHAHABUDDIN P. Asymptotic optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9(2):117–152, 1999.

[GHS99b] GLASSERMAN P., HEIDELBERGER P., and SHAHABUDDIN P. Importance sampling in the Heath-Jarrow-Morton framework. *Journal of Derivatives*, 9:32–50, 1999.

[HES93] HESTON S.L. 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.

[HH80] HALL P. and HEYDE C. *Martingale limit theory and its applications.* Academic press, 1980.

[HUL03] HULL J.C. *Options, Futures, and Other Derivatives.* Prentice Hall, Fifth Edition, 2003.

[HW87] HULL J. and WHITE A. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987.

[KS88] KARATZAS I. and SHREVE S.E. *Brownian Motion and Stoshastic Calculus.* Springer-Verlag, 1988.

[KY03] KUSHNER H. J. and YIN G. G. *Stochastic Approximations and Recursive Algorithms and Applications.* Springer, 2003.

[LL61] LASALLE J.P. and LEFSCHETZ S. *Stability by Liapounov's Direct Method with Applications.* Academic Press, New York, 1961.

[LL96] LAMBERTON D. and LAPEYRE B. *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall, 1996.

[LPS98] LAPEYRE B., PARDOUX E., and SENTIS R. *Méthodes de Monte Carlo pour les équations de transport et de diffusion.* Springer, 1998.

[MAD02] MADRAS N. *Lecture on Monte Carlo methods.* Fields Institute Monograph, AMS, 2002.

[MD02] MARTOLI N. and DEL MORAL P. *Simulation et algorithmes stochastiques.* Cépaduès-éditions, 2002.

[MR97] MUSIELA M. and RUTHKOWSKI M. *Martingale Method in Financial Modelling.* Springer, 1997.

[NEV72] NEVEU J. *Martingales à temps discrèt.* Masson, 1972.

[NEW94] NEWTON N.J. Variance reduction for simulated diffusions. *SIAM J. Appl. Math.*, 54(6):1780–1805, 1994.

[PAG03] PAGES G. Sur quelques algorithmes récursifs pour les probabiltés numériques. *Journées MAS 2000, Rennes*, 8-11 sept 2003.

[PEL96] PELLETIER M. *Sur le comportement asymptotique des algorithmes stochastiques.* PhD Thesis, Université Paris-Sud, 1996.

[PFL96] PFLUG G.CH. *Optimization of stochastic models.* Kluwer Academic Publisher, 1996.

[POL90] POLYAK B.T. New stochastic approximation type procedures. *Automatica i Telemekh (in Russian –translated as Autom. & Remote Contr., 51(7))*, pages 98–107, 1990.

[PRO90] PROTTER P. *Stochastic Integration and Differential Equations, A New Approach.* Springer-Verlag, 1990.

[REI93] REIDER R. An efficient monte carlo technique for pricing options. *Working Paper, Warton School, University of Pennsylvania*, 1993.

[RM98] RUBISTEIN R.Y. and MELAMED B. *Modern Simulation and Modeling.* John Wiley & Sons, NY, 1998.

[ROG95] ROGERS L.C.G. Which model for the term structure of interest rates should one use? *Mathematical Finance*, IMA 65:93–116, 1995.

[RUP88] RUPPERT D. Efficient estimators from a slowly convergent robbins-monro process. *Technical report, School of Oper. Res. and Ind. Eng.,Cornell Univ. Ithaca, N.Y.*, N. 781, 1988.

[SF02] SU Y. and FU M.C. Optimal importance sampling in securities pricing. *Journal of Computational Finance*, 5(4):27–50, 2002.

[SH00] SHAPIRO A. and HOMEM-DE-MELLO T. On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM, J. Optimatization*, 11(1):70–86, 2000.

[TAL90] TALAY D. and TUBARO. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8, 1990.

[VD98] VAZQUEZ-ABAD F. and DUFRESN D. Accelerated simulation for pricing asian options. *Proceedings of the Winter Simulation Conference*, pages 1493–1500, 1998.

[WIL95] WILLIAMS D. *Probability with Martingales*. Cambridge University Press, 1995.

Résumé: Dans cette thèse, nous proposons de nouvelles techniques de réduction de variance, pour les simulations Monte Carlo. Par un simple changement de variable, nous modifions la loi de simulation de façon paramétrique. L'idée consiste ensuite à utiliser une version convenablement projetée des algorithmes de Robbins-Monro pour déterminer le paramètre optimal qui "minimise" la variance de l'estimation. Nous avons d'abord développé une implémentation séquentielle dans laquelle le paramètre optimal est estimé asymptôtiquement puis injecté dans la simulation Monte Carlo. Ensuite, nous avons développé une version adaptative, dans laquelle la variance est réduite dynamiquement au cours des itérations Monte Carlo. Enfin, dans la dernière partie de notre travail, l'idée principale a été d'interprèter la réduction de variance en termes de minimisation d'entropie relative entre une mesure de probabilité optimale donnée, et une famille paramétrique de mesures de probabilité. Nous avons prouvé des résultats théoriques généraux qui définissent un cadre rigoureux d'utilisation de ces méthodes, puis nous avons effectué plusieurs expérimentations en finance et en fiabilité qui justifient de leur efficacité réelle.

Abstract: The objective of this work is to present new competitive variance reduction techniques for Monte Carlo simulations. The methods use importance sampling scheme. By an elementary change of variable, we introduce a drift term into the computation of an expectation via Monte Carlo simulations. Subsequently, the basic idea is to use a truncated version of the Robbins-Monro algorithms to find the optimal drift that reduces the variance. First, we develop a sequential application of the method, in which the optimal drift is estimated separately and is plugged in the Monte Carlo simulation. In the second part of our work we develop an adaptative version of the method, where the change of drift is selected dynamically through the Monte Carlo simulation. The last part of the work follows a similar idea but its main contribution is the introduction of a new minimisation criterion: the Kullback-Leibler entropy (or relative entropy) between two probability measures. We develop two applications of the procedure for variance reduction in Monte Carlo computation in finance and in reliability.

Discipline: PROBABILITÉS APPLIQUÉES.

MOTS-CLÉS: Méthode de Monte Carlo, réduction de variance, Algorithmes Stochastiques, Projection fixe, Projection aléatoire "à la Chen", Martingales, Entropie relative.

ADRESSE DU LABORATOIRE
CERMICS
École Nationale des Ponts et Chaussées
6 et 8 Avenue Blaise Pascal, Cité Descartes
77455 Marne-La-Vallée cedex 2.