

THESE

pour obtenir le titre de

Docteur en Sciences

de l'UNIVERSITE de Nice-Sophia Antipolis

Discipline : Mathématiques Appliquées

présentée et soutenue par

Olivier BARDOU

**Contrôle dynamique des erreurs
de simulation et d'estimation
de processus de diffusion**

Thèse dirigée par Denis TALAY
et préparée à l'INRIA Sophia Antipolis, Projet OMEGA,
soutenue le 22 mars 2005.

Jury :

M	Pierre Del Moral	Professeur	Examineur
Mme	Valentine Genon-Catalot	Professeur	Examineur
M	Peter Imkeller	Professeur	Rapporteur
M	Gilles Pagès	Professeur	Rapporteur
M	Bernard Roynette	Professeur	Examineur
M	Denis Talay	DR INRIA	Examineur

Remerciements

Mes premiers remerciements vont aux Professeurs Gilles Pagès et Peter Imkeller qui ont accepté de rapporter cette thèse. Leurs commentaires ont permis de clarifier de nombreux points des démonstrations. Je remercie d'autant plus Gilles Pagès car c'est à lui que je dois mon intérêt pour les algorithmes stochastiques.

Mes plus vifs remerciements vont à Denis Talay pour m'avoir guidé dans mes travaux de recherche tout au long de ces trois dernières années. Je ne saurais dire tout ce que lui doit cette thèse. J'espère que nos collaborations futures en seront le reflet.

Je remercie également les Professeurs Pierre Del Moral, Valentine Genon-Catalot et Bernard Roynette pour avoir accepté de participer au jury de cette thèse. Nombre de leurs travaux respectifs m'ont maintes fois inspiré au cours de la rédaction de ce travail.

Il serait sans doute trop long de remercier tous ceux que j'ai pu croiser au cours de ces années à Sophia Antipolis. Je me dois néanmoins de citer Miguel Martinez et Bouhari Arouna, Arturo Kohatsu-Higa pour m'avoir accueilli à Barcelone ainsi que les membres (présents et passés) du projet OMEGA : Mireille Bossy, Etienne Tanré, Christophe Berthelot, Awa Diop, Marie-Line Ramfos et Dalila Mana. J'ai une pensée particulière pour Axel Grorud dont j'ai pu apprécier, trop brièvement, la grande culture scientifique et la profonde humanité.

Parallèlement à cette thèse, j'ai eu la chance de participer à une étude menée par le projet OMEGA pour la Direction de la Recherche de la société Gaz de France. Cette collaboration a grandement contribué à enrichir mon travail de recherche au cours de ces trois années et je remercie vivement les agents de Gaz de France qui l'ont rendue possible : Solenne Gueydan, Gregory Benmenzer, Jeanne Rey et Marion Lacombe.

Et, bien sûr, cette thèse n'aurait pu voir le jour sans le soutien quotidien de tous mes proches, et en particulier de Catherine et Dorian. Merci à Chris sans qui je n'aurais peut-être même pas débuté cette aventure et à mes parents, pour tout.

*A Catherine et Dorian,
A mes parents.*

Table des matières

Remerciements	iii
Introduction	vii
1. Techniques de réduction de variance pour les méthodes de Monte-Carlo	vii
2. Contributions à la statistique des processus de diffusion	ix
3. Collaboration industrielle	x
partie 1. Techniques de réduction de variance pour la simulation de diffusions	1
Chapitre 1. EFFICIENT VARIANCE REDUCTION FOR FUNCTIONALS OF DIFFUSIONS BY RELATIVE ENTROPY MINIMIZATION	3
1. Introduction	3
2. The Importance Sampling paradigm	3
3. Variance reduction as entropy minimization	6
4. The Variance Reduction Procedure	8
5. Numerical Results	12
6. Conclusion	16
Chapitre 2. SIMULATIONS ERGODIQUES EFFICACES	19
1. Motivations	19
2. Représentation de Poisson et théorèmes asymptotiques classiques	20
3. Estimation dynamique de la variance	28
4. Techniques de réduction de variance	38
5. Conclusion et perspectives	50
partie 2. Deux contributions à la Statistique des diffusions	51
Chapitre 3. DÉVELOPPEMENTS D'EDGEWORTH POUR LES FONCTIONNELLES DE PROCESSUS DE DIFFUSION ERGODIQUES	53
1. Introduction	53
2. Développement d'Edgeworth pour les martingales locales continues	59
3. Développement d'Edgeworth pour les martingales locales continues normalisées	65
4. Développement d'Edgeworth pour les moyennes empiriques	74
5. Développements de Cornish-Fisher	83
6. Conclusion	86
Chapitre 4. STATISTICAL ESTIMATION OF SKEW PROCESSES	87
1. Introduction	87

2. Estimation of α when γ is known	88
3. Estimation of the parameters α and γ when they are both unknown	97
4. Numerical results	108
5. Conclusion	112
partie 3. Collaboration industrielle	113
Chapitre 5. MODELISATION ET SIMULATION DES INDICES PETROLIERS	115
1. Introduction	115
2. La construction d'un modèle pour les indices pétroliers	116
3. La simulation numérique du modèle	126
Bibliographie	129

Introduction

Cette thèse se compose de trois parties. Bien qu'indépendantes, elles reposent toutes trois sur l'étude du comportement en temps long de processus aléatoires. Plus précisément, nous étudions dans différents contextes, des méthodes adaptatives qui permettent de contrôler dynamiquement l'erreur de simulation (ou d'estimation suivant le cas) de processus de diffusion. La première partie concerne l'étude de méthodes de réduction de variance pour les simulations de type Monte-Carlo. La deuxième partie traite deux thèmes de statistique des processus. Enfin, la troisième partie synthétise une étude menée au sein du projet OMEGA pour la Direction de la Recherche de la société Gaz de France.

1. Techniques de réduction de variance pour les méthodes de Monte-Carlo

Le premier sujet abordé au cours de ce travail est l'étude de techniques de réduction de variance. Deux problématiques sont présentées. La première concerne une méthode de réduction de variance adaptative pour les simulations Monte-Carlo et l'application de cette méthode en finance. Dans un second chapitre, nous étudions les moyens de développer des méthodes de réduction de variance dans un cadre ergodique.

1.1. Technique de réduction de variance adaptative. Pour estimer l'espérance inconnue $\mathbb{E}[X]$ d'une variable aléatoire X simulable, une méthode très simple consiste à l'approcher par la moyenne empirique $\frac{1}{N} \sum_{i=1}^N X_i$ où les X_i sont des répliques indépendantes et identiquement distribuées de la variable X . C'est le principe des méthodes de Monte-Carlo. Sous des hypothèses très faibles sur la variable X (typiquement X de carré intégrable), il est connu (loi des grands nombres) que cette méthode fournit une estimation consistante de $\mathbb{E}[X]$:

$$\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow[N \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[X].$$

Par ailleurs, il est possible de caractériser l'erreur aléatoire de cette simulation grâce au théorème de la limite centrale :

$$\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}[X] \right) \xrightarrow[N \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma^2)$$

où σ^2 représente la variance de X . Ce dernier résultat fait apparaître que, pour obtenir un intervalle de confiance de taille 2ε donnée, la durée de la simulation devrait être proportionnelle à $\frac{\sqrt{N}}{\varepsilon\sigma}$. Pour obtenir une précision donnée sans trop accroître le nombre N de simulations, il est naturel d'envisager des techniques de réduction de variance. Il s'agit en général de construire une variable aléatoire Y de même espérance que X mais de variance plus faible. Or, si de nombreuses techniques sont connues dans le cadre des méthodes de Monte-Carlo

traditionnelles (variable de contrôle, fonction d'importance, intégration par parties entre autres), leur mise en œuvre requière toujours la connaissance de quantité dont l'estimation est plus difficile que la simulation de $\mathbb{E}[X]$ elle-même.

Dans ce travail, réalisé en collaboration avec Bouhari Arouna de l'ENPC, ce problème est résolu de manière algorithmique. Poursuivant les travaux antérieurs de Bouhari Arouna, nous considérons la technique bien connue de la fonction d'importance. Sa mise en œuvre repose sur un changement de probabilité; la probabilité optimale (celle qui minimise la variance) étant bien entendu inconnue. L'idée fondamentale est d'exprimer cette probabilité optimale comme la solution d'un problème de minimisation. Ce problème est ensuite résolu au cours de la simulation à l'aide d'un algorithme de type Robbins-Monro. Dans cet algorithme, les étapes ne sont plus indépendantes. Mais il est encore possible d'établir des résultats de type loi des grands nombres et théorèmes de la limite centrale qui assurent la validité de notre démarche.

Dans cet article, nous avons retenu la finance comme champ d'application. La variable X est alors une fonctionnelle de processus de diffusion et le changement de probabilité repose sur le théorème de Girsanov. Dans ce contexte, nous montrons que notre algorithme fournit également une estimation du delta; la sensibilité par rapport à la condition initiale.

Cet algorithme a également été appliqué avec succès dans le cadre de la collaboration avec GDF détaillée dans la troisième partie de ce mémoire.

1.2. Simulation ergodiques efficaces. Ce second chapitre traite du cas particulier où l'espérance que nous cherchons à estimer s'exprime sous la forme $\int f d\mu$ où la mesure μ admet une densité p solution de l'équation de Fokker-Planck stationnaire $\mathcal{L}^*p = 0$. Dans ce cas, un estimateur naturel de $\int f d\mu$ est la moyenne empirique $\frac{1}{T} \int_0^T f(X_s) ds$ du processus de diffusion associé au générateur infinitésimal \mathcal{L} . Cet estimateur vérifie des propriétés de consistance et de normalité asymptotique similaires à l'estimateur de Monte-Carlo traditionnel. Toutefois, outre les problèmes de réduction de variance, une autre difficulté caractérise ce genre de simulations. En effet, sa variance admet une expression explicite $\sigma_f^2 = -2 \int f G_f d\mu$ qui fait intervenir la solution du problème de Poisson $\mathcal{L}G_f = -(f - \int f d\mu)$. La résolution de ce problème est évidemment très coûteuse et requière en outre la connaissance de $\int f d\mu$. Jusqu'à présent, il était donc très difficile d'obtenir une estimation de la variance.

Le premier objectif de ce chapitre est donc d'introduire un estimateur simple de cette variance. Notre démarche repose sur des résultats de convergence fonctionnelle tel que le théorème de la limite centrale presque sûr. Plus précisément, nous commençons par montrer à l'aide des résultats de Pardoux et Veretennikov sur l'équation de Poisson, que, pour toutes fonction f mesurable et à croissance polynomiale, les moyennes empiriques admettent une représentation martingale de la forme

$$\int_0^T f(X_s) ds - \int f d\mu = \int_0^T (\sigma G'_f)(X_s) dW_s + G_f(x) - G_f(X_T)$$

où σ est le coefficient de diffusion du processus X . Ce résultat nous permet d'étendre le théorème de la limite centrale presque sûr aux moyennes empiriques. Précisément, nous montrons que pour toute fonction φ à croissance au plus quadratique,

$$\frac{1}{\ln T} \int_1^T \varphi \left(\frac{1}{\sqrt{t}} \int_0^t \left(f(X_s) - \int f d\mu \right) ds \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \int \varphi d\mathcal{N}(0, \sigma_f^2)$$

et nous donnons une vitesse de convergence associée. Un corollaire immédiat nous fournit un estimateur de la variance asymptotique. Des résultats similaires sont en outre établis pour le temps local d'un processus de diffusion.

Dans une seconde partie, nous introduisons trois méthodes de réduction de variance pour les simulations ergodiques. Il semble qu'il n'existe pas de travaux antérieurs sur la question. Deux de ces techniques sont inspirées de méthodes connues dans le cadre iid (la variable de contrôle et l'intégration par parties). Une troisième méthode, fondée sur un changement de temps, est particulière à notre contexte. Dans chaque cas, nous établissons des résultats de convergence et nous nous efforçons de caractériser le domaine d'application de ces différentes techniques.

2. Contributions à la statistique des processus de diffusion

La deuxième partie de ce mémoire est consacrée à deux problèmes de statistique des processus. Dans un premier chapitre, nous nous sommes intéressés à des résultats précisant les théorèmes de la limite centrale usuels. Nous avons ainsi établi des développements de type Edgeworth pour différentes variables souvent rencontrées en statistique des processus. Le second chapitre de cette partie introduit un problème nouveau : l'estimation statistique des processus de diffusion à dérive singulière ou processus de diffusion asymétrisés. Nous présentons en particulier une méthode d'estimation efficace dans le cas unidimensionnel.

2.1. Statistiques d'ordre supérieur pour les processus de diffusion. Bien qu'il ne concerne pas la statistique des processus, le chapitre deux a permis d'introduire des outils souvent rencontrés en statistique : la loi des grands nombres et le théorème de la limite centrale pour les martingales et les moyennes empiriques de processus de diffusion. Dans ce chapitre nous cherchons à raffiner ces résultats classiques de convergence ainsi que leurs analogues pour les martingales normalisées par leur crochet. Cette dernière quantité intervient en particulier lors de l'estimation par maximum de vraisemblance.

En nous fondant sur le développement asymptotique des moments de martingales et sur l'argument de Feller dans le cas iid, nous avons pu établir des développements d'Edgeworth à l'ordre 1 pour les trois quantités mentionnées précédemment. Ce résultat est établi dans un premier temps pour les martingales puis étendu aux martingales normalisées et aux moyennes empiriques par la méthode du delta. Certains de ces résultats sont déjà connus, mais notre méthode de preuve permet d'affaiblir les hypothèses supposées dans les travaux antérieurs. Cette méthode ouvre en outre la perspective d'une extension de ces résultats aux fonctionnelles additives martingales de processus ergodiques plus généraux.

Plusieurs applications de ces résultats sont étudiés, autant en statistique qu'en simulation. En particulier, nous établissons des développements de Cornish-Fisher pour les quantiles dans la perspective du bootstrap des intervalles de confiance.

2.2. Statistique des processus asymétrisés. Ce quatrième chapitre, écrit en collaboration avec Miguel Martinez (INRIA), introduit une problématique nouvelle ; l'étude statistique des processus asymétrisés. Nous appelons ainsi un processus X solution de l'EDS

$$X_t = x + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s + (2\alpha - 1)L_t^\gamma(X) - K_t^{[a, b]}(X), \alpha \in]0, 1[.$$

Le terme $K_t^{[a,b]}$ induit une réflexion du processus sur l'intervalle $[a, b]$. De tels processus généralisent le mouvement brownien asymétrique et interviennent lors de la modélisation de systèmes physiques en milieux discontinus. Il n'existe à notre connaissance aucun résultat concernant l'estimation de leurs paramètres α et γ .

Nous proposons dans cet article une méthodologie pour estimer les coefficients du modèles dans une asymptotique de temps long. Pour ce faire, nous commençons par établir un théorème ergodique pour de tels processus. L'estimation du paramètre α s'effectue par une méthode de moments généralisée et celle de γ par la minimisation d'un contraste bien choisi. Dans les deux cas, l'ergodicité du processus nous permet d'obtenir la consistance forte de ces estimateurs et nous donnons de plus une vitesse de convergence pour l'estimateur du paramètre α .

3. Collaboration industrielle

Cette dernière partie synthétise les résultats d'une collaboration entre le projet OMEGA et la société GDF. Cette collaboration a duré trois ans et comportait deux volets. Dans un premier temps, il s'agissait de construire un modèle valide pour l'évolution de différents indices pétroliers. La seconde étape consistait en la mise en œuvre d'une méthode de Monte-Carlo efficace pour la simulation de fonctionnelles complexes de ces indices.

Les données fournies par GDF consistaient en dix séries de données journalières s'étendant sur 10 ans. Pour les modéliser, nous avons opté pour un modèle de diffusions. Le coefficient de diffusion du processus fût déterminé par une procédure non-paramétrique à l'aide d'estimateurs à noyaux de type Nadaraya-Watson. Du fait de la structure des données, il semblait raisonnable de considérer une dérive linéaire dont les coefficients furent estimés par maximum de vraisemblance. Au cours de ce travail, nous fûmes amenés à considérer un problème semble-t-il nouveau : la détermination du nombre de mouvements browniens sous-jacents au modèle. La question précise est la suivante : quelle est la dimension minimale du mouvement brownien sous-jacent nécessaire pour rendre compte de la corrélation entre les différents indices ?

La deuxième partie de cette collaboration portait sur la simulation de ce modèle. Etant donné la dimension du problème et la complexité des fonctionnelles à évaluer, introduire une technique de réduction de variance s'imposait. Nous avons donc mis en œuvre la méthode de réduction de variance adaptative décrite dans le premier chapitre de ce mémoire. Différents tests numériques ont pu mettre en évidence son efficacité sur ce problème.

Première partie

Techniques de réduction de variance pour la
simulation de diffusions

Cette première partie consiste en deux chapitres.

- Le premier chapitre présente une technique de réduction de variance adaptative reposant sur la méthode traditionnelle de fonction d'importance. Après avoir rappelé le principe de cette méthode, nous montrons qu'elle est équivalente à un problème de minimisation faisant intervenir la distance de Kullback-Leibler. Nous décrivons ensuite comment la solution de ce problème peut être approchée itérativement au cours de la simulation grâce à un algorithme de Robbins-Monro. Différentes applications en finance illustrent notre méthodologie.
- Le second chapitre aborde les méthodes de simulation ergodiques du point de vue de la réduction de variance. Nous commençons par rappeler les fondements de ces méthodes et comment caractériser la variance de telles simulations. La variance en question s'exprime à l'aide de la solution d'un problème de Poisson et de l'espérance à estimer. La question de son estimation n'est donc pas triviale. Nous montrons qu'il est possible d'obtenir un estimateur consistant de cette variance grâce à un théorème de la limite centrale presque sûr pour les moyennes empiriques. Ce chapitre se conclue sur la présentation de trois techniques de réduction de variance dans le contexte ergodique.

EFFICIENT VARIANCE REDUCTION FOR FUNCTIONALS OF DIFFUSIONS BY RELATIVE ENTROPY MINIMIZATION

En collaboration avec Bouhari Arouna (ENPC)

1. Introduction

Examples of successful using of Importance Sampling applied to derivatives pricing are Reider [76], Glasserman, Heidelberger and Shahabuddin [38] or Boyle, Broadie and Glasserman [17]. More closely related to this work are the papers by Su and Fu [82], Vasquez-Abad and Dufresne [86] and Arouna [3] 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 [2], such an idea was combined with the Robbins-Monro algorithm to develop an Adaptive 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 [71] 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 discrete 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 minimization 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 efficiency of the method is numerically tested on a wide range of derivatives including path-dependent options and interest rate caps.

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 $\mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ defined on this space. We consider

the \mathbb{R}^n valued process X , solution to the stochastic differential equation

$$(1) \quad \begin{cases} X_0 &= x \\ dX_t &= b(X_t) dt + \sigma(X_t) dW_t, t \in [0, T]. \end{cases}$$

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, $\forall x, y \in \mathbb{R}^n$,

$$\begin{aligned} \|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| &\leq K\|x - y\|, \\ \|b(x)\| + \|\sigma(x)\| &\leq C\|x\| \end{aligned}$$

such that the SDE (1) has a unique strong solution (see Karatzas and Shreve [47]). 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

$$\mathbb{P}(\psi_T(X) \geq \varepsilon) = 1.$$

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

$$(2) \quad \mathcal{E}^T := \left\{ \theta, \mathbb{F}\text{-adapted such that } \mathbb{E} \left[\int_0^T \|\theta_t\|^2 dt \right] < +\infty \text{ and } \mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right) \right] < +\infty \right\}$$

and

$$(3) \quad \mathcal{H}^T := \left\{ \theta, \mathbb{F}\text{-adapted such that } \mathbb{E} \left[\int_0^T \|\theta_t\|^2 dt \right] < +\infty \right\}.$$

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 [75]).

THEOREM 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}_θ 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 2.0.2. *Given a probability measure \mathbb{P}_θ , equivalent to \mathbb{P} on (Ω, \mathcal{F}_T) , there exists an \mathbb{F} -adapted process $(\theta_t)_{t \leq T}$ such that $\int_0^T \|\theta_t\|^2 dt < +\infty$ \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}_θ on (Ω, \mathcal{F}_T) , the process X is solution to the following SDE

$$(4) \quad \begin{cases} X_0 &= x \\ dX_t &= (b(X_t) + \sigma(X_t)\theta_t) dt + \sigma(X_t) dW_t^\theta, t \in [0, T]. \end{cases}$$

Moreover, we have

$$(5) \quad \mathbb{E} [\psi_T(X)] = \mathbb{E}_\theta \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right],$$

where the subscript indicates that the expectation is computed with respect to \mathbb{P}_θ . It follows that, under the probability \mathbb{P}_θ , $\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta}$ 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 $\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta}$ under \mathbb{P}_θ *i.e.* the process θ that solves

$$(6) \quad \min_{\theta \in \mathcal{E}^T} \text{Var}_\theta \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right].$$

Theoretically, it is known that there exists a unique probability measure \mathbb{P}_{θ^*} for which $\text{Var}_{\theta^*} \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_{\theta^*}} \right]$ vanishes. In fact, since

$$\begin{aligned} \text{Var}_\theta \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right] &= \mathbb{E}_\theta \left[\left(\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right)^2 \right] - \mathbb{E}_\theta \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right]^2 \\ &= \mathbb{E} \left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right] - \mathbb{E} [\psi_T(X)]^2, \end{aligned}$$

\mathbb{P}_{θ^*} should satisfy

$$\mathbb{E} \left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_{\theta^*}} \right] = \mathbb{E} [\psi_T(X)]^2,$$

and this relation leads obviously to

$$\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}} = \frac{\psi_T(X)}{\mathbb{E} [\psi_T(X)]},$$

which is equivalent to

$$(7) \quad d\mathbb{P}_{\theta^*} = \frac{\psi_T(X)}{V_0} d\mathbb{P}.$$

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 build an efficient estimator of the process θ^* . This problem will be formulated as a distance minimization issue. We are going to see that a clever choice of the distance is important from a numerical point of view. The next section is thus devoted to the comparison between different distances.

3. Variance reduction as entropy minimization

3.1. χ^2 minimization. The χ^2 divergence between two probability measures is formally defined by

$$\begin{aligned}\mathcal{R}_2(\mathbb{P}, \mathbb{Q}) &= \frac{1}{2} \int \frac{(d\mathbb{P} - d\mathbb{Q})^2}{d\mathbb{Q}} \text{ if } \mathbb{P} \sim \mathbb{Q} \\ &= +\infty \text{ otherwise.}\end{aligned}$$

When considering this distance, simple computations show that problem (6) is equivalent to the following distance minimization problem between \mathbb{P}_θ and \mathbb{P}_{θ^*}

$$(8) \quad \min_{\theta \in \mathcal{E}^T} \mathcal{R}_2(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta).$$

To see this, first observe that

$$\begin{aligned}\mathcal{R}_2(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta) &= \frac{1}{2} \int \frac{(d\mathbb{P}_{\theta^*} - d\mathbb{P}_\theta)^2}{d\mathbb{P}_\theta} \\ &= \frac{1}{2} \left(\int \frac{(d\mathbb{P}_{\theta^*})^2}{d\mathbb{P}_\theta} + 1 \right).\end{aligned}$$

Plugging equation (7) in this expression gives

$$\begin{aligned}\mathcal{R}_2(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta) &= \frac{1}{2} \left(\int \frac{\psi_T(X)^2}{\mathbb{E}[\psi_T(X)]^2} \frac{d\mathbb{P}}{d\mathbb{P}_\theta} d\mathbb{P} + 1 \right) \\ &= \frac{1}{2} \left(\frac{\mathbb{E} \left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right]}{V_0^2} + 1 \right).\end{aligned}$$

Therefore, problem (8) reexpresses as

$$(9) \quad \min_{\theta \in \mathcal{E}^T} \mathbb{E} \left[\psi_T(X)^2 \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right].$$

By the Girsanov Theorem this latter problem reduces to

$$(10) \quad \min_{\theta \in \mathcal{E}^T} \mathbb{E} \left[\psi_T(X)^2 \exp \left(- \int_0^T \theta_t \cdot dW_t + \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right) \right].$$

Equation (9) is exactly the usual criterion to minimize when referring to optimal importance sampling! See for instance Newton [71] and Arouna [3]. From this last 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 clever choice of the distance can overcome this difficulty.

3.2. Kullback-Leibler minimization. This distance is defined by

$$\begin{aligned}\mathcal{R}_0(\mathbb{P}, \mathbb{Q}) &= \int \ln \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{P} \text{ if } \mathbb{P} \sim \mathbb{Q} \\ &= +\infty \text{ otherwise.}\end{aligned}$$

Our aim is to minimize $\mathcal{R}_0(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta)$. This quantity can be rewritten as

$$\begin{aligned}\mathcal{R}_0(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta) &= \int \ln \left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_\theta} \right) d\mathbb{P}_{\theta^*} \\ &= \int \ln \left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}} \cdot \frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) d\mathbb{P}_{\theta^*} \\ &= \int \ln \left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}} \right) d\mathbb{P}_{\theta^*} + \int \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) d\mathbb{P}_{\theta^*}.\end{aligned}$$

The first integral does not depend on \mathbb{P}_θ . Then, minimizing $\mathcal{R}_0(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta)$ is equivalent to minimizing $\int \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) d\mathbb{P}_{\theta^*}$. Moreover, since

$$d\mathbb{P}_{\theta^*} = \frac{\psi_T(X)}{\mathbb{E}[\psi_T(X)]} d\mathbb{P},$$

we get,

$$\begin{aligned}\int \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) d\mathbb{P}_{\theta^*} &= \int \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) \frac{\psi_T(X)}{\mathbb{E}[\psi_T(X)]} d\mathbb{P} \\ &= \frac{\mathbb{E} \left[\psi_T(X) \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) \right]}{\mathbb{E}[\psi_T(X)]}.\end{aligned}$$

Thus, the problem of minimizing $\mathcal{R}_0(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta)$ reduces to

$$(11) \quad \min_{\theta \in \mathcal{E}^T} \mathbb{E} \left[\psi_T(X) \ln \left(\frac{d\mathbb{P}}{d\mathbb{P}_\theta} \right) \right],$$

which we rewrite as, by the virtue of the Girsanov Theorem,

$$(12) \quad \min_{\theta \in \mathcal{E}^T} J(\theta),$$

with

$$(13) \quad J(\theta) := \mathbb{E} \left[\psi_T(X) \left(- \int_0^T \theta_t \cdot dW_t + \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right) \right].$$

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. Moreover, the Kullback Leibler distance is sharper than the χ^2 divergence in the sense that

$$(14) \quad \mathcal{R}_0(\mathbb{P}, \mathbb{Q}) \leq \mathcal{R}_2(\mathbb{P}, \mathbb{Q}), \quad \forall \mathbb{P} \sim \mathbb{Q},$$

since $\ln x \leq \frac{1}{2}(x^2 - 1)$, $\forall x > 0$. Finally, the most important fact is the quadratic dependance of the right hand side of equation (12) instead of the exponential dependance of equation (10) upon θ . Clearly, (12) leads to a more tractable minimization procedure than (10). We refer to Basseville [9] for complements on entropy and pseudo-distances between probability measures.

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 [71] that the zero-variance drift term θ^* of an option with discounted payoff ψ_T is given by

$$\theta_t^* = \frac{\sigma(X_t)\pi_t}{\mathbb{E}[\psi_T(X)|\mathcal{F}_t]}, \quad t \in [0, T]$$

where the process π_t is the *unique* process implicitly defined by the Martingale Representation theorem

$$(15) \quad \psi_T(X) = \mathbb{E}[\psi_T(X)] + \int_0^T \pi_t \cdot \sigma(X_t) dW_t, \quad a.s.$$

The process π is referred 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 :

$$(16) \quad \pi_t = \mathbb{E}[\psi_T(X)|\mathcal{F}_t] \sigma^{-1}(X_t) \cdot \theta_t^* \quad a.s.$$

which reduces to

$$(17) \quad \Delta := V_0 \sigma^{-1}(x) \cdot \theta_0^*,$$

for $t = 0$. We now develop the discrete framework to perform the computation of θ^* and V_0 .

4. The Variance Reduction Procedure

In practice, when the solution to the SDE (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

$$\hat{V}_0 = \mathbb{E}[\psi_T(\bar{X}^h)],$$

where \bar{X}^h is a discretization of the solution X to (1). For sake of simplicity, we shall focus on the Euler-Maruyama scheme (see Talay and Tubaro [85]) although our method could be extended to any other scheme. The discretization of equation (1) with constant time step $h := \frac{T}{m}$ is then

$$\begin{cases} \bar{X}_0 & = X_0 \\ \bar{X}_{h(p+1)} & = \bar{X}_{hp} + b(\bar{X}_{hp})h + \sigma(\bar{X}_{hp})\sqrt{h}\Delta W_p, \quad p = 0 \dots m-1 \end{cases}$$

with $\Delta W_p \stackrel{iid}{\sim} \mathbb{N}(0, I_d)$ and I_d the identity matrix of \mathbb{R}^d . Similarly, equation (4) is discretized by the following scheme :

$$\begin{cases} \tilde{X}_0 & = X_0 \\ \tilde{X}_{h(p+1)} & = \tilde{X}_{hp} + \left(b(\tilde{X}_{hp}) + \sigma(\tilde{X}_{hp}) \tilde{\theta}_p \right) h + \sigma(\tilde{X}_{hp})\sqrt{h}\Delta W_p, \quad p = 0 \dots m-1. \end{cases}$$

In this representation, $\tilde{\theta} := (\tilde{\theta}_0, \dots, \tilde{\theta}_{m-1})$ is a $d \times m$ matrix. Let us define the interpolated process $(\tilde{\theta}_t)_{t \leq T}$ by setting

$$\tilde{\theta}_t := \tilde{\theta}_p, \quad \text{if } t \in [hp, h(p+1)[.$$

Then $\tilde{\theta}$ can be seen as an approximation of the random vector process θ . In the sequel ΔW denotes the $d \times m$ matrix $(\Delta W_0, \dots, \Delta W_{m-1})$. In this discrete setting, equation (5) translates into

$$(18) \quad \begin{aligned} \hat{V}_0 &= \mathbb{E} [\psi_T(\bar{X})] \\ &= \mathbb{E} \left[\psi_T(\tilde{X}) \exp \left(-\tilde{\theta} \cdot \sqrt{h} \Delta W - \frac{h}{2} \|\tilde{\theta}\|^2 \right) \right] \end{aligned}$$

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 (12), the Kullback-Leibler minimization becomes

$$(19) \quad \min_{\tilde{\theta} \in \mathcal{E}^T} \tilde{J}(\tilde{\theta}), \quad \tilde{\theta} \in \mathbb{R}^{d \times m},$$

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

$$(20) \quad \nabla \tilde{J}(\tilde{\theta}) = 0, \quad \tilde{\theta} \in \mathbb{R}^{d \times m},$$

with

$$\nabla \tilde{J}(\tilde{\theta}) := \mathbb{E} \left[\psi_T(\bar{X}) \left(h\tilde{\theta} - \sqrt{h} \Delta W \right) \right].$$

4.1. Efficiency of a Kullback-Leibler based variance reduction procedure. 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 the process $\tilde{\theta}$. Thus, this process 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 $\|\tilde{\theta}\| \rightarrow +\infty$, 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 the process $\tilde{\theta}$ (see Arouna [1]). Moreover, remember that the χ^2 distance is always greater than the Kullback-Leibler distance (see (14)). Thus, the relation linking these two distances is illustrated in figure 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.

4.2. The Martingale Monte Carlo Algorithm. Following Arouna [2], 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 (18) and a result of Arouna [2] we have

$$(21) \quad \tilde{V}_N := \frac{1}{N} \sum_{i=1}^N \psi_T(\tilde{X}^i) \exp \left(-\tilde{\theta}^i \cdot \sqrt{h} \Delta W^i - \frac{h}{2} \|\tilde{\theta}^i\|^2 \right) \xrightarrow[N \rightarrow +\infty]{a.s.} \mathbb{E} [\psi_T(\bar{X})],$$

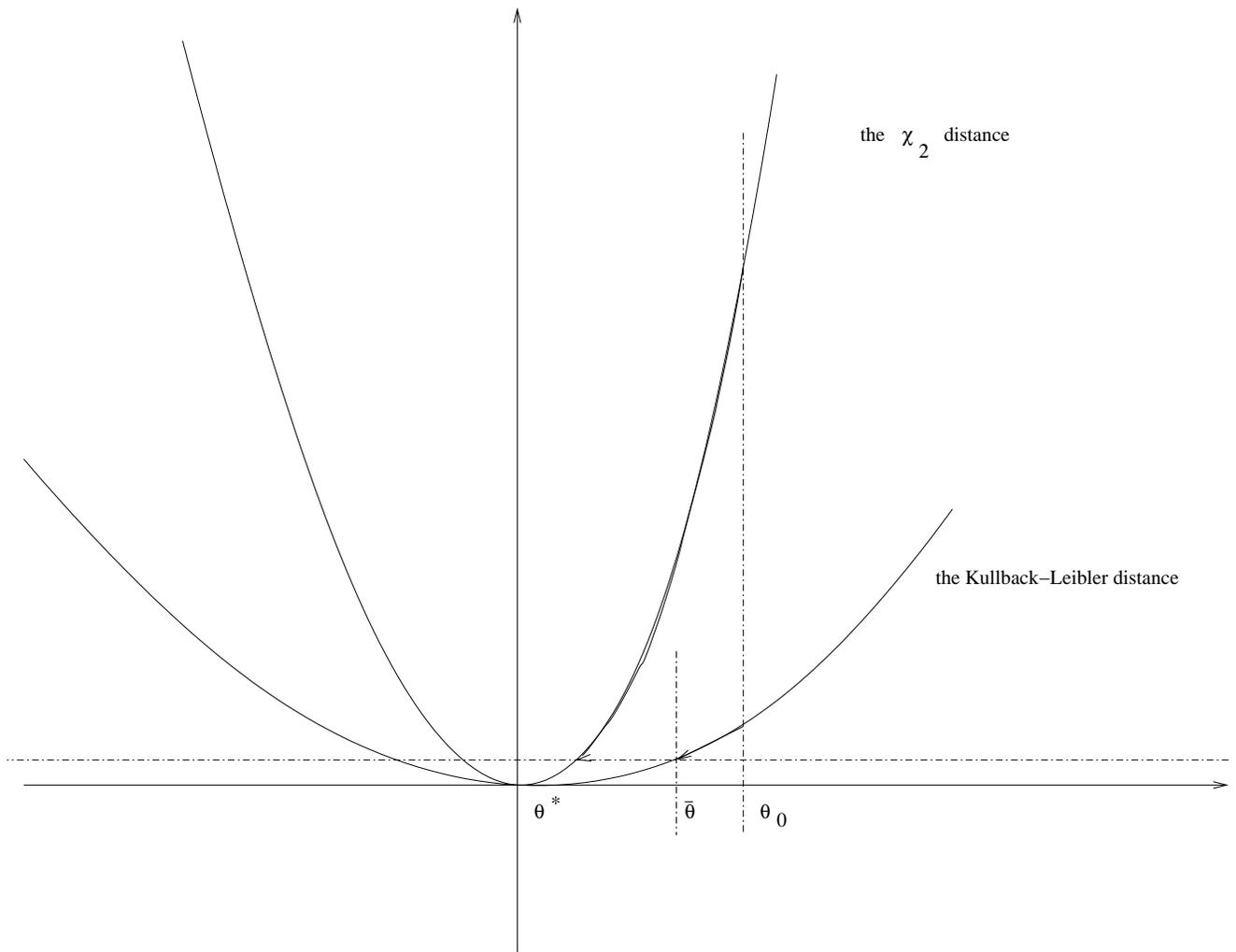


FIG. 1. Variance and cross-entropy minimization

where $(\tilde{X}^i)_{i=1\dots N}$ 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

$$(22) \quad \Sigma_N^2 := \frac{1}{N} \sum_{i=1}^N \psi_T^2(\tilde{X}^i) \exp\left(-2\tilde{\theta}^i \cdot \sqrt{h}\Delta W^i - h\|\tilde{\theta}^i\|^2\right) - \tilde{V}_N^2,$$

and it is proved in Arouna [2] that, on the one hand,

$$(23) \quad \Sigma_N^2 \xrightarrow[N \rightarrow +\infty]{a.s.} \text{Var}_{\tilde{\theta}^*} \left[\psi_T(X) \frac{d\mathbb{P}}{d\mathbb{P}_{\tilde{\theta}^*}} \right]$$

and, on the other hand,

$$(24) \quad \sqrt{N} \frac{\tilde{V}_N - \mathbb{E}[\psi_T(\tilde{X})]}{\Sigma_N^2} \xrightarrow[N \rightarrow +\infty]{\text{Law}} \mathcal{N}(0, 1).$$

Thus, we have a central limit theorem that enables us to built confidence intervals. Moreover the asymptotic variance of the estimation (21) is clearly smaller than that 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 (17) by setting

$$(25) \quad \hat{\Delta} := \tilde{V}_N \sigma^{-1}(0, x) \cdot \tilde{\theta}_0^N.$$

Next the minimization issue (20) will be performed using a simple Robbins-Monro procedure as described below.

4.3. 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 $\Delta W \sim \mathbb{N}(0, I_{d \times m})$. 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

$$(H_0) : \exists \alpha > 0 \quad \text{such that} \quad \mathbb{E}[\varphi_T^{2+\alpha}(\Delta W)] < +\infty.$$

Our approach for solving problem (20) relies on the following iterative scheme

$$(26) \quad \tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma_{i+1} F(\tilde{\theta}^i, \Delta W^i),$$

where $(\gamma_i)_{i \geq 1}$ is a sequence of positive numbers converging to 0. This iterative stochastic approximation is known as Robbins–Monro algorithm and is well indicated for the resolution of the zero-search problem

$$(27) \quad \mathbb{E}[F(\tilde{\theta}^*, \Delta W)] = 0, \quad \tilde{\theta}^* \in \mathbb{R}^{d \times m}.$$

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 $(\tilde{\mathcal{F}}_i)_{i=1 \dots n}$ the σ -algebra generated by the random vectors $(\Delta W^i)_{i=1 \dots n}$. 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

$$(28) \quad \mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2 | \tilde{\mathcal{F}}_{i-1}] \leq C(1 + \|\tilde{\theta}^i\|^2), \quad a.s.$$

where

$$C = \max(\mathbb{E}[\varphi_T^2(\Delta W)], \mathbb{E}[\varphi_T^2(\Delta W) \|\Delta W\|^2])$$

is a finite positive constant by the virtue of hypothesis (H_0) . Next, we state a convergence Theorem for (26) proved in Dufflo [26] :

THEOREM 4.3.1. *Under the following hypotheses*

$$(29) \quad \exists! \tilde{\theta}^* \in \mathbb{R}^{d \times m} \text{ such that } E(\tilde{\theta}^*) = 0 \text{ and } \forall \tilde{\theta} \in \mathbb{R}^{d \times m} \text{ with } \tilde{\theta} \neq \tilde{\theta}^*, (\tilde{\theta} - \tilde{\theta}^*) \cdot E(\tilde{\theta}) > 0,$$

$$(30) \quad \sum_i \gamma_i = +\infty \quad \text{and} \quad \sum_i \gamma_i^2 < +\infty,$$

$$(31) \quad \mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2 | \tilde{\mathcal{F}}_{i-1}] < K(1 + \|\tilde{\theta}^i\|^2), \text{ a.s.},$$

the sequence of random vectors $(\tilde{\theta}^i)_{i \geq 0}$ defined in (26) 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 (29) is a simple consequence of the Theorem of finite increments. Since hypothesis (31) is fulfilled from equation (28), Theorem (4.3.1) applies for the iterative algorithm (26).

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 address this issue in the next section.

4.4. Summary. Finally, let us summarize the Martingale Monte Carlo algorithm to be implemented in the very general case.

$$\tilde{\theta}^1 = \theta^0, \quad h = \frac{T}{p}, \quad \gamma^i \xrightarrow{i \rightarrow +\infty} 0, \quad \sum_{i=1}^{+\infty} \gamma^i = +\infty.$$

for $i = 1 \dots m$,

$$\bar{X}_0^i = X_0, \quad \tilde{X}_0^i = X_0, \quad \Delta W^i \sim \mathcal{N}(0, I_{d,p})$$

for $j = 1 \dots p$,

$$\bar{X}_j^i = \bar{X}_{j-1}^i + b(\bar{X}_{j-1}^i)h + \sigma(\bar{X}_{j-1}^i) \cdot \sqrt{h} \Delta W_j^i$$

$$\tilde{X}_j^i = \tilde{X}_{j-1}^i + \left(b(\tilde{X}_{j-1}^i) + \sigma(\tilde{X}_{j-1}^i) \cdot \tilde{\theta}_{j-1}^i \right) h + \sigma(\tilde{X}_{j-1}^i) \cdot \sqrt{h} \Delta W_j^i$$

$$\tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma^{i+1} \psi_T(\bar{X}^i) \left(\sqrt{h} \tilde{\theta}^i - \Delta W^i \right)$$

$$\tilde{V}_N := \frac{1}{N} \sum_{i=1}^N \psi_T(\tilde{X}^i) \exp \left(-\sqrt{h} \text{Tr} \left(\tilde{\theta}^i \cdot {}^t \Delta W^i \right) - \frac{h}{2} \|\tilde{\theta}^i\|_F^2 \right)$$

Of course, the Robbins-Monro algorithm and the Monte-Carlo procedure can also be performed separately.

5. Numerical Results

5.1. Preliminary remarks.

5.1.1. *Choice of the sequence γ_i .* It is worth pointing out that the Robbins–Monro algorithm (26) 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 recommendations for financial applications that are made at section 6 of Arouna [3]. For a more global insight into this problem, one can see Kushner and Yin [53], where the optimal numerical convergence of the Robbins–Monro algorithm is discussed in length.

5.1.2. *Reduction of the conditional variance.* Remember that the observation of the the Robbins-Monro algorithm (26) 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 $\mathbb{E}[\|F(\tilde{\theta}^i, \Delta W^i)\|^2 | \mathcal{F}_i]$ of the sequence of the observations. This can be done by replacing in equation (26) 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 (4.3.1) still applies for the Robbins-Monro algorithm defined as

$$(32) \quad \tilde{\theta}^{i+1} = \tilde{\theta}^i - \gamma_{i+1} \hat{F}(\tilde{\theta}^i, \Delta W^i),$$

5.1.3. *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

$$(33) \quad \tilde{\theta}^{i+1} = \begin{cases} \tilde{\theta}^i - \gamma_{i+1} \hat{F}(\tilde{\theta}^i, \Delta W^i) & \text{if } \left\| \tilde{\theta}^i - \gamma_{i+1} \hat{F}(\tilde{\theta}^i, \Delta W^i) \right\| \leq U, \\ \tilde{\theta}^i & \text{otherwise.} \end{cases}$$

U is a constant upperbound of $\tilde{\theta}^*$ taken equal to 5 in our all experimentations.

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

$$(34) \quad \bar{X}_{t_i} = \bar{X}_{t_{i-1}} e^{(r-\sigma^2/2)(t_i-t_{i-1}) + \sigma\sqrt{t_i-t_{i-1}}\Delta W_i}, \quad \bar{X}_0 = 1, \quad t_i = \frac{i}{n}, \quad i = 1 \dots n.$$

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

$$\text{RV} := \frac{\frac{1}{N} \sum_{i=1}^N \psi_T^2(\bar{X}^i) - \left(\frac{1}{N} \sum_{n=1}^N \psi_T(\bar{X}^i) \right)^2}{\Sigma_N^2},$$

where Σ_i^2 is defined in (22). The numerical results for this case are in Table 1 and show the effectiveness of the procedure. **PrixRM**, **PrixBS**, Δ_{RM} and Δ_{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 (25)) 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

α	Parameters			RM+IS			
	σ	K/S_0	PrixRM	Δ_{RM}	PrixBS	Δ_{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 : $\bar{X}_0 = 50$, $r = 0.05$, $T = 1.0$.

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 (34). Numerical procedures are available for the pricing of this option with continuous averaging but not with discret averaging (see Geman and Yor [36]). In this latter case, simulation is necessary. The numerical results obtained are summarized in Table 2. Δ_{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 portfolio computed with our algorithm.

Table 2 Estimated variance reduction ratio and delta hedging for an arithmetic asian call option in the Black–Scholes model

n	Parameters			RM+IS			
	α	σ	K/S ₀	PrixRM	Δ_{RM}	Δ_{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 : $\bar{X}_0 = 50$, $r = 0.05$, $T = 1.0$.

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

$$\frac{d\bar{X}_t^i}{\bar{X}_t^i} = rdt + \sum_{j=1}^d \sigma_{ij} dW_t^j, \quad \bar{X}_0^i > 0,$$

with $W = (W^1, \dots, W^d)$ a d -dimensional Brownian motion 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 $\sqrt{RV} > 3$. 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

$$\frac{d\bar{X}_t}{\bar{X}_t} = rdt + \sigma\sqrt{d} dB_t, \quad \bar{X}_0 > 0,$$

with $\sigma = \sigma_{ij}$, $\forall i, j$ 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 $\Delta_{BS} = 0.69$ share of each underlying asset. With these values as inputs, we obtained the following hedging portofolio

$$\Delta_{RM} = \{0.73, 0.69, 0.67, 0.69, 0.70, 0.71, 0.70, 0.72, 0.70, 0.72\},$$

after $RM = 20,000$ iterations of the Robbins-Monro algorithm (33). 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

n	Parameters			RM+IS	
	α	σ	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
20	25		1.4	1.79	23.3
	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 : $\bar{X}_0 = 50$, $r = 0.05$, $T = 1.0$.

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.

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[71]). 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.

SIMULATIONS ERGODIQUES EFFICACES

1. Motivations

L'objet de ce chapitre est l'estimation efficace de quantités du type $\int f d\mu$ où f est une fonction connue et μ , l'unique mesure invariante du processus de diffusion X de \mathbb{R}^n , solution de l'EDS

$$(35) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0$$

où W est un mouvement brownien standard sur \mathbb{R}^d . Nous noterons \mathbb{P}_x , la probabilité sous laquelle $X_0 = x$, presque sûrement. Sous des hypothèses de régularité sur les fonctions $b : \mathbb{R}^n \mapsto \mathbb{R}^n$ et $\sigma : \mathbb{R}^n \mapsto \mathcal{M}_{n \times d}(\mathbb{R})$, la mesure μ admet une densité p régulière, solution de l'équation de Fokker-Planck stationnaire :

$$(36) \quad \frac{1}{2} \sum_{i,j=1}^n \frac{\partial}{\partial x_j} \left(a_{ij} \frac{\partial p}{\partial x_i} \right) + \sum_{i=1}^n \frac{\partial}{\partial x_i} (\bar{b}_i p) = 0$$

avec $a := \sigma\sigma^*$ et $\bar{b}_i := \frac{1}{2} \sum_{j=1}^n \frac{\partial a_{ij}}{\partial x_j} - b_i$, $i = 1 \dots n$. Cette équation s'écrit également

$$\mathcal{L}^* p = 0$$

où \mathcal{L}^* désigne l'adjoint formel du générateur infinitésimal du processus X :

$$(37) \quad \mathcal{L} := \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i \frac{\partial}{\partial x_i}.$$

L'étude de la solution de cette équation en dimension grande est essentielle dans la résolution de nombreux problèmes issus de la physique ou des sciences du vivant. La mécanique aléatoire (voir l'ouvrage de C. Soize [80]) et la magnétoencéphalographie (voir la thèse récente de M. Martinez [67]) sont deux exemples de tels problèmes.

Pour répondre à la question de l'estimation de $\int f d\mu$, nous considérons une méthode probabiliste reposant sur le

THÉORÈME 1.0.1 (Théorème ergodique). *Si le processus X , solution de l'EDS (35), est ergodique, de mesure de probabilité invariante unique μ alors, pour toute fonction $f \in L^1(\mu)$,*

$$\frac{1}{T} \int_0^T f(X_t) dt \xrightarrow[T \rightarrow +\infty]{\mathbb{P}_x\text{-p.s.}} \int f d\mu, \quad \mu(dx)\text{-p.s.}$$

Nous renvoyons à l'ouvrage de R. Z. Has'minskii [42] pour une preuve de ce résultat. En pratique, l'implémentation numérique de cette idée s'appuie sur un schéma de discrétisation

$(\bar{X}_p)_{p \geq 0}$ du processus X et sur la simulation de la moyenne empirique

$$\frac{1}{N} \sum_{p=1}^N f(\bar{X}_p).$$

Deux type d'erreur sont donc à considérer :

- une erreur de discrétisation $\int f d\mu - \int f d\bar{\mu}$ provenant de l'approximation de la mesure μ par la mesure invariante $\bar{\mu}$ de la chaîne de Markov $(\bar{X}_p)_{p \geq 0}$ et reflétant la précision de la simulation,
- une erreur statistique $\frac{1}{N} \sum_{p=1}^N f(\bar{X}_p) - \int f d\bar{\mu}$ reflétant la vitesse de convergence de l'algorithme.

C'est cette deuxième erreur que nous nous proposons d'étudier au cours de ce chapitre. Pour bénéficier des outils puissants du cadre continu, nous négligerons l'erreur de discrétisation en précisant aux moments opportuns dans quelle mesure elle affecte nos résultats.

Dans un premier temps, nous rappellerons un certain nombre de résultats classiques concernant les deux erreurs évoquées précédemment. En particulier, nous caractériserons l'erreur statistique à l'aide d'un théorème de la limite centrale. Nous présenterons ensuite une méthode adaptative permettant d'estimer cette erreur au cours de la simulation ergodique. Enfin, nous exposerons trois techniques de réduction de variance permettant d'améliorer l'efficacité de la simulation ergodique.

2. Représentation de Poisson et théorèmes asymptotiques classiques

Cette première partie est consacrée au rappel de résultats classiques concernant la convergence des processus de diffusion. Après avoir présenté nos hypothèses de travail, nous introduirons notre outil principal : la solution de l'équation de Poisson associée à l'opérateur \mathcal{L} défini en (37). Diverses applications des propriétés de cette solution seront présentées ainsi qu'un panorama des résultats existants concernant les schémas de discrétisation.

2.1. Hypothèses. La notion sous-jacente à l'ensemble de cette étude est la convergence des semi-groupes de diffusion vers leur éventuelle mesure invariante. Parmi la vaste littérature sur le sujet, nous ferons principalement référence aux travaux de H. Ganidis, B. Roynette et F. Simonot [35], D. Talay [83], et E. Pardoux et A. Yu. Veretennikov [72]. Sauf mention contraire, ce sont sous les hypothèses de ces derniers que nous nous placerons pour toute la suite de ce chapitre. Celles-ci sont les suivantes.

(H1) La fonction a est bornée, uniformément continue et uniformément elliptique :

$$\exists K > 0 / \forall x, \xi \in \mathbb{R}^n, \sum_{i,j} a_{i,j}(x) \xi_i \xi_j \geq K |\xi|^2.$$

(H2) La fonction b est localement bornée et il existe $\alpha > -1$, $M > 0$ et $r > 0$ tels que

$$b(x) \cdot \frac{x}{|x|} \leq -r|x|^\alpha, |x| \geq M.$$

(H3) La fonction f n'est pas constante et il existe $\beta \geq 0$, $C_1 \geq 0$ et $C_2 \geq 0$ tels que

$$|f(x)| \leq C_1 + C_2|x|^\beta.$$

REMARQUE. Les hypothèses de [72] sont en fait plus générales car elles autorisent $\alpha = -1$. Les preuves que nous allons présenter peuvent également s'adapter à ce cas mais deviennent plus techniques. Nous ne le considérerons donc pas explicitement.

Une première conséquence de notre jeu d'hypothèses est le

THÉORÈME 2.1.1. *Sous les hypothèses (H1)–(H2), le processus X défini en (35) admet une unique mesure invariante de probabilité μ et toute fonction f satisfaisant (H3) appartient à $L^1(\mu)$.*

DÉMONSTRATION. Nous renvoyons à [72] pour une preuve de ce résultat. □

REMARQUE. Ces hypothèses ne garantissent pas forcément l'existence d'une densité p pour la mesure μ . Mais celle-ci n'est pas nécessaire pour établir nos résultats. Si cette densité existe et est de classe C^2 , alors il est connu qu'elle est solution de l'équation de Fokker-Planck stationnaire (36).

Nous pouvons alors définir

$$\tilde{f} := f - \int f d\mu.$$

2.2. Equation de Poisson. Nous nous intéressons maintenant à l'équation de Poisson

$$(38) \quad \mathcal{L}G_f = -\tilde{f}.$$

où \mathcal{L} est l'opérateur défini en (37). Le résultat suivant est également dû à E. Pardoux et A. Yu. Veretennikov [72].

THÉORÈME 2.2.1. *Sous les hypothèses (H1)–(H3), la fonction G_f définie par*

$$(39) \quad G_f(x) := \int_0^{+\infty} \mathbb{E}_x [\tilde{f}(X_t)] dt, \quad x \in \mathbb{R}^n$$

est continue, appartient à l'espace de Sobolev $W_{p,loc}^2$ pour tout $p > 1$, est solution de l'équation (38) et satisfait les trois propriétés suivantes. Pour tout $m > \beta \vee 2$, il existe des constantes C_m et C'_m telles que

$$(40) \quad |G_f(x)| \leq C_m (1 + |x|^m), \quad x \in \mathbb{R}^n$$

et

$$(41) \quad |\nabla G_f(x)| \leq C'_m (1 + |x|^m), \quad x \in \mathbb{R}^n.$$

En particulier, G_f , ∇G_f et leurs puissances sont dans $L^1(\mu)$. De plus,

$$(42) \quad \int G_f d\mu = 0$$

et G_f est l'unique solution, dans $W_{p,loc}^2$ pour tout $p > 1$, de l'équation (38) qui vérifie les propriétés (40) et (42).

REMARQUE. En dimension 1, l'équation de Fokker-Planck (36) et l'équation de Poisson (38) peuvent se résoudre explicitement. En effet, sous l'hypothèse

$$(43) \quad \int_{\mathbb{R}} \frac{\exp(2 \int_0^x \frac{b(z)}{\sigma^2(z)} dz)}{\sigma^2(x)} dx < \infty,$$

nous avons l'existence d'une densité p pour la mesure invariante. Et cette densité admet une expression explicite (voir Gihman et Skorohod [37]).

$$(44) \quad p(x) \propto \frac{\exp(2 \int_0^x \frac{b(z)}{\sigma^2(z)} dz)}{\sigma^2(x)}.$$

Il est alors facile de vérifier que

$$G_f(x) = -2 \int_{\tilde{x}}^x \frac{1}{\sigma^2(y)p(y)} \int_{-\infty}^y \tilde{f}(z)p(z) dz dy$$

où \tilde{x} est choisi de telle sorte que (42) soit vérifiée.

2.3. Représentation de Poisson et applications. Cette nouvelle partie est dévolue à l'outil fondamental de notre étude.

2.3.1. *Représentation de Poisson.* Le lemme fondamental est le suivant.

LEMME 2.3.1. *Sous les hypothèses (H1)–(H3),*

$$(45) \quad \int_0^T \tilde{f}(X_s) ds = \int_0^T (\sigma \nabla G_f)(X_s) dW_s - (G_f(X_T) - G_f(x))$$

\mathbb{P}_x -presque sûrement pour tout $x \in \mathbb{R}^n$.

DÉMONSTRATION. Comme les coefficients b et σ sont localement bornés et que la fonction a est non-dégénérée, cette représentation est une application directe de la formule d'Itô-Krylov (voir Krylov [51] pour un énoncé dans le cas d'un domaine borné et Pardoux et Veretennikov [72] pour l'extension dans tout l'espace) à la fonction G_f dont nous savons par le théorème 2.2.1 qu'elle appartient à $W_{p,loc}^2$ pour tout $p > 1$. \square

Ce lemme est fondamental pour l'étude de la convergence des fonctionnelles additives de processus ergodiques. L'équation de Poisson est un outil très utilisé pour l'étude des processus ergodiques dans un cadre discret (voir par exemple Meyn et Tweedie [69]). Nous renvoyons aux articles de R. Liptser, V. Spokoiny et A. Yu. Veretennikov ([61], [62], [63]) ou encore E. Fournié [32] pour d'autres applications dans le cadre continu. Nous commençons par rappeler comment cette représentation permet d'obtenir la vitesse de convergence du théorème ergodique 1.0.1.

REMARQUE. Sous nos hypothèses, un résultat de Pardoux et Veretennikov [72] nous assure que, pour tout $m > 0$, il existe une constante $C_m > 0$ indépendante de t et telle que

$$\mathbb{E}_x [|X_t|^m] < C_m.$$

Ainsi, nous avons, pour tout $T > 0$,

$$\mathbb{E}_x \left[\int_0^T |\sigma \nabla G_f|^2(X_s) ds \right] < \infty.$$

Par conséquent, $\int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s$ est une vraie martingale de carré intégrable.

2.3.2. *Théorème de la limite centrale.* Nous commençons par définir une quantité qui sera fondamentale dans la suite de ce travail.

LEMME 2.3.2. Soit

$$(46) \quad \sigma_f^2 := \int |\sigma \nabla G_f|^2 d\mu.$$

Sous les hypothèses (H1)–(H3),

$$0 < \sigma_f^2 < +\infty.$$

DÉMONSTRATION. La fonction σ est supposée bornée et nous savons (voir théorème 2.2.1) que $\nabla G_f \in L^2(\mu)$. Par conséquent $\sigma_f^2 < +\infty$. D'autre part, il est clair, comme la mesure μ n'est pas dégénérée, que σ_f^2 ne peut être nulle que si la fonction G_f est constante. Ceci ne peut survenir que si la fonction f est elle-même constante, ce qui est contraire à nos hypothèses. Sur ce dernier point, nous renvoyons également à Bhattacharya [13]. \square

REMARQUE. Une intégration par parties et l'utilisation de l'équation de Fokker-Planck (36) conduisent à l'expression alternative de σ_f^2 :

$$\sigma_f^2 = -2 \int f G_f d\mu.$$

Voir également Bhattacharya [13].

La proposition suivante caractérise l'erreur statistique de la simulation ergodique.

PROPOSITION 2.3.1. *Sous les hypothèses (H1)–(H3),*

$$(47) \quad \frac{1}{\sqrt{T}} \int_0^T \tilde{f}(X_s) ds \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_f^2).$$

DÉMONSTRATION. D'après la représentation de Poisson (45),

$$\frac{1}{\sqrt{T}} \int_0^T \tilde{f}(X_s) ds = \frac{1}{\sqrt{T}} \int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s - \frac{1}{\sqrt{T}} (G_f(X_T) - G_f(x)).$$

Or, nous vérifions que

$$\begin{aligned} \frac{1}{T} \left\langle \int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s \right\rangle_T &= \frac{1}{T} \int_0^T |\sigma \nabla G_f|^2(X_s) ds \\ &\xrightarrow[T \rightarrow +\infty]{\mathbb{P}_{x\text{-p.s.}}} \sigma_f^2, \forall x \in \mathbb{R}^n \end{aligned}$$

d'après le théorème ergodique 1.0.1. Par ailleurs, le lemme 2.3.2 nous assure que $0 < \sigma_f^2 < +\infty$. Nous pouvons alors appliquer le théorème de la limite centrale pour les martingales continues (voir Kutoyants [54]) et conclure que

$$(48) \quad \frac{1}{\sqrt{T}} \int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_f^2).$$

Par ailleurs, comme X_T converge en loi et que la fonction G_f est continue, $G_f(X_T)$ converge également en loi et le lemme de Slutsky nous permet alors d'affirmer que

$$(49) \quad \frac{1}{\sqrt{T}} (G_f(X_T) - G_f(x_0)) \xrightarrow[T \rightarrow +\infty]{\mathbb{P}} 0.$$

Une nouvelle application du lemme de Slutsky combinant les équations (48) et (49) conclut la preuve. \square

Un intervalle de confiance asymptotique de niveau α pour l'espérance $\int f d\mu$ est donné par

$$\left[\frac{1}{T} \int_0^T f(X_s) ds - \frac{\sigma_f q_\alpha}{\sqrt{T}}, \frac{1}{T} \int_0^T f(X_s) ds + \frac{\sigma_f q_\alpha}{\sqrt{T}} \right]$$

avec q_α , le quantile de niveau $\alpha/2$ de la loi gaussienne standard. Il apparaît clairement que la précision du calcul est contrôlée par le rapport $\frac{\sigma_f}{\sqrt{T}}$. Concrètement, pour obtenir une précision ε avec probabilité α , il faudrait choisir une durée de simulation

$$T \simeq \frac{\sigma_f^2 q_\alpha^2}{\varepsilon^2}.$$

De ce fait, la précision désirée sera atteinte d'autant plus rapidement que la variance σ_f^2 sera petite. Deux questions se posent alors. D'une part, comment estimer efficacement cette variance σ_f^2 dont l'expression théorique est impraticable? D'autre part, est-il possible de mettre en œuvre des procédures permettant de réduire cette variance? Nous fournissons dans les deux dernières sections de ce chapitre des éléments de réponse à ces questions.

Auparavant, nous revenons brièvement sur le choix de nos hypothèses et montrons comment généraliser les résultats précédents en dimension un.

2.3.3. *Discussion des hypothèses.* Les hypothèses retenues par D. Talay [83] sont un cas particulier des hypothèses précédentes où les coefficients du modèle et la fonction f sont supposés très réguliers et $\alpha = 1$ (hypothèses de Has'minskiï). L'auteur démontre alors une convergence exponentielle vers la mesure invariante ainsi que des estimations sur le semi-groupe du processus qui impliquent entre autre l'existence et la régularité de la solution du problème de Poisson (38). La démonstration de la représentation de Poisson ne repose alors que sur la formule d'Itô classique.

H. Ganidis, B. Roynette et F. Simonot [35] se placent dans le cas d'un coefficient de dérive de type gradient et étudient en détails la vitesse de convergence vers la mesure invariante en fonction du paramètre α . En général la fonction f est supposée bornée, sauf dans le cas $\alpha > 1$ où l'hypercontractivité du semi-groupe permet de relaxer cette hypothèse.

2.3.4. *Généralisation en dimension un.* Le cadre établi précédemment ne permet pas de traiter le problème de l'estimation de la densité invariante elle-même. Il faudrait en effet pouvoir établir l'analogie des résultats énoncés jusqu'ici pour le cas où f n'est définie qu'au sens des distributions. Ceci est possible en dimension un en faisant appel aux propriétés du temps local. Le cadre unidimensionnel n'est évidemment que de peu d'intérêt dans la perspective de simulations numériques. Nous mentionnons toutefois cette extension par souci d'être complet. Nous supposons que l'hypothèse (43) est vérifiée et que la mesure invariante admet une densité dont l'expression est donnée par (44). Nous supposons de plus que la fonction σ est dérivable pour garantir la dérivabilité de cette densité.

Densité empirique. En dimension un, la mesure empirique du processus X au point $x \in \mathbb{R}$ est définie par

$$\mu_T(x) := \frac{1}{T} \int_0^T \mathbb{1}_{(X_s \leq x)} ds.$$

La formule des temps d'occupation permet de lui donner l'expression alternative

$$\begin{aligned} \mu_T(x) &= \frac{1}{T} \int_{\mathbb{R}} \frac{\mathbb{1}_{(y \leq x)} L_T^y}{\sigma^2(y)} dy \\ &= \frac{1}{T} \int_{-\infty}^x \frac{L_T^y}{\sigma^2(y)} dy \end{aligned}$$

où $(L_t^y)_{t \geq 0}$ désigne le temps local du processus X au point y . Il est alors naturel de définir la densité empirique du processus X au point x par

$$p_T(x) := \frac{L_T^x}{T \sigma^2(x)}.$$

Il est connu que, sous des hypothèses assez faibles, cet estimateur est fortement consistant (voir Bosq et Davydov [16] et Van Zanten[88]). Par ailleurs, un résultat de Kutoyants [56] fournit une représentation de Poisson dans ce contexte :

$$(50) \quad p_T(x) - p(x) = \frac{1}{T} \int_0^T \sigma(X_t) \dot{G}_x(X_t) dW_t - \frac{1}{T} (G_x(X_T) - G_x(X_0))$$

avec $G_x \in L^1(\mu)$, une primitive de la fonction

$$(51) \quad y \mapsto -2p(x) \frac{\mathbb{1}_{(y>x)} - \mu(y)}{\sigma^2(y)p(y)}$$

telle que

$$(52) \quad \int G_x d\mu = 0.$$

Notons que cette fonction est dérivable, mais pas C^1 . Mentionnons les travaux de A. Dalalyan [25] pour des application de cette formule en statistique des processus. Nous allons maintenant vérifier que la dénomination “représentation de Poisson” n’est pas abusive en prouvant que cette fonction est bien solution faible du problème de Poisson associé, c’est-à-dire que $G_x(\cdot)$ est une fonction de Green associée à l’opérateur \mathcal{L} . Soit φ , une fonction de classe C^∞ à support compact. Nous avons

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{L}G_x(y)\varphi(y) dy &= \int_{\mathbb{R}} \left(b\dot{G}_x + \frac{\sigma^2}{2}\ddot{G}_x \right) (y)\varphi(y) dy \text{ par définition de } \mathcal{L} \\ &= \int_{\mathbb{R}} \dot{G}_x(y) \left((b - \sigma\dot{\sigma})\varphi - \frac{\sigma^2}{2}\dot{\varphi} \right) (y) dy \text{ après intégration par parties} \\ &= \int_{\mathbb{R}} \left[\frac{\sigma^2}{2}\dot{G}_x \left(\varphi \frac{\dot{p}}{p} - \dot{\varphi} \right) \right] (y) dy \text{ car } \frac{\dot{p}}{p} = 2\frac{b - \sigma\dot{\sigma}}{\sigma^2} \text{ d'après (44)} \\ &= p(x) \int_{\mathbb{R}} \frac{\mu(y) - \mathbb{1}_{(y>x)}}{p(y)} \left(\varphi \frac{\dot{p}}{p} - \dot{\varphi} \right) (y) dy \text{ en utilisant (51)} \\ &= p(x) \int_{\mathbb{R}} \varphi(y) dy - \varphi(x). \end{aligned}$$

Nous avons bien vérifié que, formellement,

$$(53) \quad \mathcal{L}G_x = -(\delta_x - p(x)).$$

Variance de la simulation. Un corollaire immédiat de la représentation de Poisson (50) est le théorème de la limite centrale suivant :

$$\sqrt{T}(p_T(x) - p(x)) \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_x^2)$$

avec comme variance de la simulation

$$\begin{aligned} \sigma_x^2 &:= 4p(x)^2 \int_{\mathbb{R}} \frac{(\mathbb{1}_{(y>x)} - \mu(y))^2}{\sigma^2(y)p(y)} dy \\ &= \int (\sigma\dot{G}_x)^2 d\mu. \end{aligned}$$

Ou encore, après intégration par parties et en utilisant (53) et (52),

$$(54) \quad \sigma_x^2 = -2p(x)G_x(x).$$

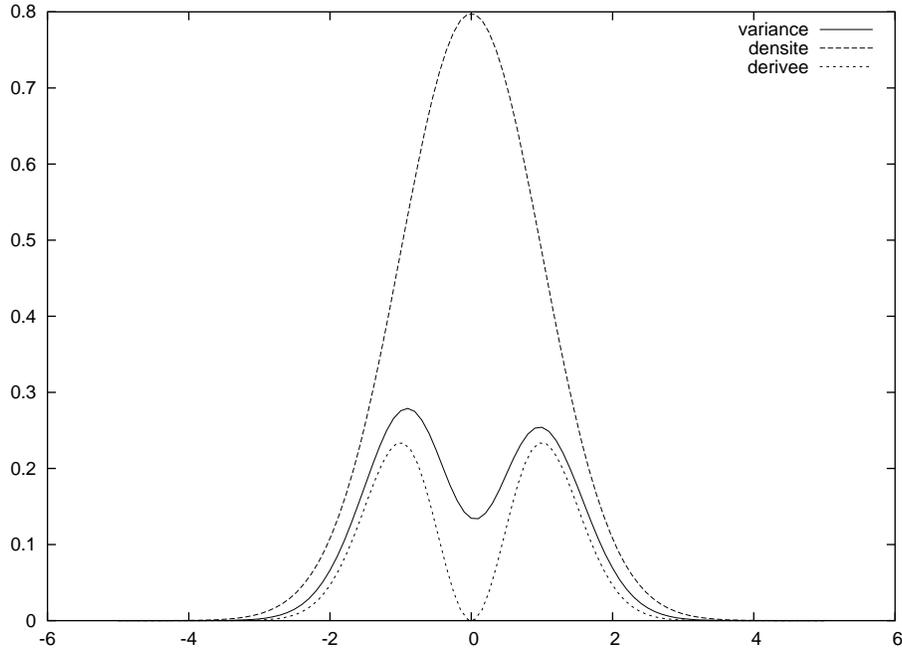


FIG. 1. Variance ergodique $x \mapsto \sigma_x^2$, densité $x \mapsto p(x)$ et dérivée de la densité au carré $x \mapsto \dot{p}(x)^2$ pour le processus d'Ornstein-Uhlenbeck de paramètre $\theta = \frac{1}{2}$.

Contrairement au cadre des méthodes de Monte-Carlo traditionnelles, l'estimation de la densité par simulation ergodique est une simulation de variance finie. Par conséquent, il est théoriquement possible d'estimer la densité en tout point avec une précision arbitraire, sans faire exploser le temps de calcul. Remarquons également que la fonction

$$x \mapsto 4 \int_{\mathbb{R}} (\mathbb{1}_{(y>x)} - \mu(y))^2 s(y) dy$$

représente la variance relative de cette simulation.

La figure 1 représente la variance ergodique $x \mapsto \sigma_x^2$ pour le cas processus d'Ornstein-Uhlenbeck

$$X_t^y = y - \frac{1}{2} \int_0^t X_s^y ds + W_t.$$

Notons le lien avec la dérivée de la densité invariante. La figure 2 représente la variance relative de la simulation, c'est-à-dire la fonction $x \mapsto -2G_x(x)/p(x)$. Elle donne un résultat plus conforme à l'intuition ; à précision fixée, l'estimation de la densité est plus difficile dans les queues de la distribution qu'au voisinage de la moyenne.

2.4. Schémas de discrétisation. Avant de considérer le problème de l'estimation de la variance, nous présentons très brièvement quelques références concernant l'étude de l'erreur de discrétisation. Pour mettre en œuvre la méthode de simulation ergodique, il faut en pratique avoir recours à un schéma de discrétisation tel que

$$\begin{cases} \bar{X}_0 &= X_0 \\ \bar{X}_{p+1} &= \bar{X}_p + b(\bar{X}_{\eta_p})h_p + \sigma(\bar{X}_p)\sqrt{h_p}\Delta W_p, p \geq 0, \end{cases}$$

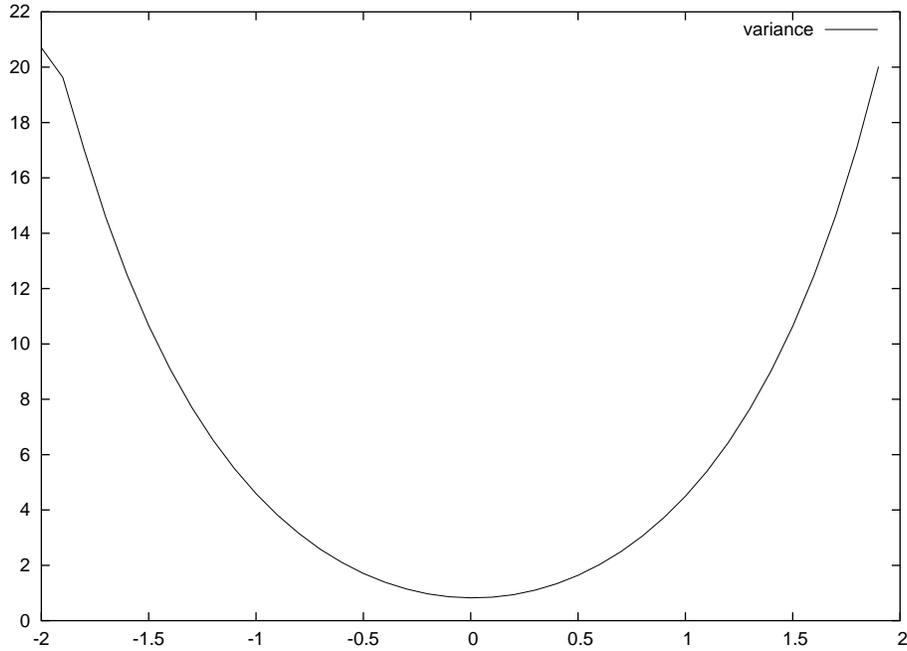


FIG. 2. Variance relative $x \mapsto \sigma_x^2/p(x)^2$ pour le processus d'Ornstein-Uhlenbeck de paramètre $\theta = \frac{1}{2}$.

avec $(\Delta W_p)_{p>0}^h$, un bruit blanc, le plus souvent gaussien. L'erreur de discrétisation est alors définie par

$$\text{p.s.-} \lim_{N \rightarrow +\infty} \left| \frac{\sum_{p=1}^N \gamma_p f(\bar{X}_p)}{\sum_{p=1}^N \gamma_p} - \int f d\mu \right|.$$

Dans le cas de coefficients b , σ et f réguliers, D. Talay et L. Tubaro [85] démontrent que l'erreur de discrétisation du schéma d'Euler-Maruyama ($\eta_p = p$, $h_p \equiv h$, $\gamma_p = 1$) se développe en puissances de h .

Si les coefficients du modèle ne sont pas globalement lipschitziens, comme cela arrive lors de la discrétisation de systèmes issus de la mécanique aléatoire (voir D. Talay [84] et J. C. Mattingly, A. M. Stuart et D. J. Higham [68]), il faut avoir recours à un schéma implicite pour s'assurer que la discrétisation conserve l'ergodicité du modèle ($\eta_p = p + 1$, $h_p \equiv h$, $\gamma_p = 1$). Dans ce cadre, D. Talay [84] démontre qu'il est encore possible d'obtenir un développement de l'erreur en puissances de h .

Signalons enfin qu'il est possible d'obtenir une erreur de discrétisation nulle en utilisant un schéma à pas décroissant ($\eta_p = p$, $h_p \rightarrow 0$, $\sum_p h_p \rightarrow +\infty$, $\sum_p \gamma_p \rightarrow +\infty$). Ces schémas sont étudiés, sous différentes hypothèses de régularité sur les coefficients du modèles par D. Lamberton et G. Pagès [58] [59].

3. Estimation dynamique de la variance

A notre connaissance, il n'existe aucun estimateur consistant de la variance asymptotique σ_f^2 dans le cadre que nous avons adopté. Nous proposons ici un tel estimateur et établissons sa

vitesse de convergence. La construction de cet estimateur repose sur un principe d'invariance fonctionnel.

3.1. Introduction. L'objet de cette section est d'établir un théorème de la limite centrale presque sûr pour la fonctionnelle

$$\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds$$

ainsi qu'une vitesse de convergence pour ce théorème. Nous commençons par rappeler l'équivalent de ces résultats pour le mouvement brownien. Les démonstrations en sont très simples et éclairent les résultats que nous énoncerons par la suite.

LEMME 3.1.1. Soit B , un mouvement brownien standard. Pour tout réel $\sigma_f > 0$, et pour toute fonction $\varphi \in L^1(\mathcal{N}(0, \sigma_f^2))$,

$$\frac{1}{\ln T} \int_1^T \varphi \left(\frac{B_{\sigma_f^2 t}}{\sqrt{t}} \right) \frac{dt}{t} \xrightarrow[T \rightarrow \infty]{\text{p.s.}} \int \varphi d\mathcal{N}(0, \sigma_f^2).$$

DÉMONSTRATION. C'est un résultat très classique dont nous rappelons la preuve. Définissons le processus $Y_t := e^{-\frac{t}{2}} B_{\sigma_f^2 e^t}$, $t \geq 0$. Ce processus vérifie l'EDS

$$dY_t = -\frac{1}{2} Y_t dt + \sigma_f d\beta_t$$

où $\beta_t := \frac{1}{\sigma_f} \int_1^{e^t} \frac{1}{\sqrt{r}} dB_{\sigma_f^2 r}$, $t \geq 0$ est un mouvement brownien standard d'après le théorème de caractérisation de Lévy. Le processus Y est donc un processus d'Ornstein-Uhlenbeck et admet pour unique mesure invariante la loi normale centrée de variance σ_f^2 . Par conséquent, pour toute fonction $\varphi \in L^1(\mathcal{N}(0, \sigma_f^2))$,

$$\frac{1}{T} \int_0^T \varphi(Y_t) dt \xrightarrow[T \rightarrow \infty]{\text{p.s.}} \int \varphi d\mathcal{N}(0, \sigma_f^2).$$

Il suffit alors de remplacer $Y_t = e^{-\frac{t}{2}} B_{\sigma_f^2 e^t}$ dans l'équation précédente puis de procéder au changement de variable $u = e^t$ pour conclure. \square

Il est également possible de préciser la vitesse de convergence du résultat précédent.

LEMME 3.1.2. Sous les hypothèses et les notations du théorème précédent, nous définissons la fonction

$$\bar{\varphi} := \varphi - \int \varphi d\mathcal{N}(0, \sigma_f^2)$$

ainsi que la solution G_φ du problème de Poisson

$$-y\dot{G}_\varphi(y) + \sigma_f^2\ddot{G}_\varphi(y) = -2\bar{\varphi}.$$

Si

$$\bar{\sigma}_\varphi^2 := -2 \int \bar{\varphi} G_\varphi d\mathcal{N}(0, \sigma_f^2) < +\infty,$$

alors

$$\frac{1}{\sqrt{\ln T}} \int_1^T \bar{\varphi} \left(\frac{B_{\sigma_f^2 t}}{\sqrt{t}} \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \bar{\sigma}_\varphi^2).$$

DÉMONSTRATION. Ce résultat se déduit du théorème 2.3.1 appliquée au processus d'Ornstein-Uhlenbeck en utilisant le même changement de variable que dans la preuve précédente. \square

Les résultats de type TCL presque sûr sont bien connus pour les suites de variables aléatoires indépendantes et identiquement distribuées. Nous nous contentons de citer le travail récent de B. Bercu [10] concernant la convergence des moments dans le TCL presque sûr. Concernant les processus, la question naturelle est de savoir si les résultats obtenus pour le mouvement brownien peuvent être étendus aux martingales. Sur ce point, ce sont les travaux de Maaouia [66] et Chaabane [22] qui sont à l'origine de nos résultats. Citons enfin l'article de D. Lambertson et G. Pagès [58] où les auteurs établissent un TCL presque sûr dans le cadre de l'estimation de la mesure invariante d'une diffusion par un schéma à pas décroissant.

3.2. Principe d'invariance avec poids logarithmique. Nous énonçons maintenant le résultat principal de cette section.

THÉORÈME 3.2.1. Soient $0 \leq \tau \leq 1$ et

$$\mathcal{G}_\tau := \left\{ \varphi \in L^2(\mathcal{N}(0, \sigma_f^2)) / |\varphi(x) - \varphi(y)| \leq K|x - y|(|x|^\tau + |y|^\tau + 1), (x, y) \in \mathbb{R}^2 \right\}$$

où K est une constante positive.

Sous les hypothèses (H1)–(H3), pour toute fonction $\varphi \in \mathcal{G}_\tau$,

$$\frac{1}{\ln T} \int_1^T \varphi \left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\mathbb{P}_x\text{-p.s.}} \int \varphi d\mathcal{N}(0, \sigma_f^2), \forall x \in \mathbb{R}^n.$$

DÉMONSTRATION. Rappelons la représentation de Poisson (45),

$$\int_0^t \tilde{f}(X_s) ds = \underbrace{\int_0^t (\sigma \nabla G_f)(X_s) \cdot dW_s}_{=: M_t} - \underbrace{(G_f(X_t) - G_f(x))}_{=: G_t}.$$

Nous avons déjà vérifié au cours de la preuve de la proposition 2.3.1 que

$$\frac{\langle M \rangle_t}{t} \xrightarrow[t \rightarrow +\infty]{\text{p.s.}} \sigma_f^2 \text{ avec } 0 < \sigma_f^2 < +\infty.$$

Sous cette condition, les travaux de Maaouia [66] et Chaabane [22] nous assurent que, pour toute fonction $\varphi \in \mathcal{G}_\tau$,

$$(55) \quad \frac{1}{\ln T} \int_1^T \varphi \left(\frac{M_t}{\sqrt{t}} \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\mathbb{P}_x\text{-p.s.}} \int \varphi d\mathcal{N}(0, \sigma_f^2), \forall x \in \mathbb{R}^n$$

Considérons maintenant la différence

$$\begin{aligned} D_T &:= \left| \frac{1}{\ln T} \int_1^T \varphi \left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right) \frac{dt}{t} - \frac{1}{\ln T} \int_1^T \varphi \left(\frac{M_t}{\sqrt{t}} \right) \frac{dt}{t} \right| \\ &\leq \frac{1}{\ln T} \int_1^T \left| \varphi \left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right) - \varphi \left(\frac{M_t}{\sqrt{t}} \right) \right| \frac{dt}{t} \\ &\leq \frac{1}{\ln T} \int_1^T \left| \varphi \left(\frac{M_t}{\sqrt{t}} - \frac{G_t}{\sqrt{t}} \right) - \varphi \left(\frac{M_t}{\sqrt{t}} \right) \right| \frac{dt}{t} \\ &\leq \frac{K}{\ln T} \int_1^T \left| \frac{G_t}{\sqrt{t}} \right| \left(\left| \frac{M_t}{\sqrt{t}} - \frac{G_t}{\sqrt{t}} \right|^\tau + \left| \frac{M_t}{\sqrt{t}} \right|^\tau + 1 \right) \frac{dt}{t} \\ &\leq \underbrace{\frac{K}{\ln T} \int_1^T \frac{|G_t|}{\sqrt{t}} \frac{|M_t - G_t|^\tau}{\sqrt{t}^\tau} \frac{dt}{t}}_{=: D_T^1} + \underbrace{\frac{K}{\ln T} \int_1^T \frac{|G_t|}{\sqrt{t}} \frac{|M_t|^\tau}{\sqrt{t}^\tau} \frac{dt}{t}}_{=: D_T^2} + \underbrace{\frac{K}{\ln T} \int_1^T \frac{|G_t|}{\sqrt{t}} \frac{dt}{t}}_{=: D_T^3}. \end{aligned}$$

Nous considérons d'abord le terme D_T^3 . Une intégration par parties nous donne

$$\begin{aligned} D_T^3 &= \frac{1}{\ln T} \left\{ \left[\frac{1}{t^{3/2}} \int_1^t |G_s| ds \right]_1^T + \frac{3}{2} \int_1^T \frac{1}{t^{5/2}} \int_1^t |G_s| ds dt \right\} \\ &= \frac{1}{\ln T} \frac{1}{T^{3/2}} \int_1^T |G_t| dt + \frac{3}{2 \ln T} \int_1^T \frac{1}{t^{3/2}} \int_1^t |G_s| ds \frac{dt}{t}. \end{aligned}$$

Rappelons que $G_t = (G_f(X_t) - G_f(x))$. Le processus X étant ergodique et comme $G_f \in L^1(\mu)$, le terme $\frac{1}{t} \int_1^t |G_s| ds$ converge presque sûrement vers une limite finie. Par conséquent,

$$\frac{1}{t^{3/2}} \int_1^t |G_s| ds \xrightarrow[t \rightarrow +\infty]{\text{p.s.}} 0.$$

Le premier terme de D_T^3 converge donc presque sûrement vers 0. Le seconde terme apparaît comme la moyenne de Césaro d'une fonction de limite nulle. Ce terme converge donc également vers 0. Finalement, nous vérifions que

$$D_T^3 \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0.$$

Pour traiter le terme D_T^2 , remarquons que l'inégalité de Cauchy-Schwarz nous donne

$$D_T^2 \leq K \left(\frac{1}{\ln T} \int_1^T \frac{|G_t|^2}{t} dt \right)^{\frac{1}{2}} \left(\frac{1}{\ln T} \int_1^T \frac{|M_t|^{2\tau}}{\sqrt{t}} \frac{dt}{t} \right)^{\frac{1}{2}}.$$

Le premier terme se traite de manière similaire à D_T^2 et converge presque sûrement vers 0. Comme, pour tout $0 \leq \tau \leq 1$, la fonction $y \mapsto |y|^{2\tau}$ appartient à \mathcal{G}_τ , nous pouvons appliquer au second terme le résultat de l'équation (55) et conclure qu'il converge vers une limite finie. Ainsi,

$$D_T^2 \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0.$$

Enfin, pour traiter l'expression D_T^1 , notons que

$$\begin{aligned} D_T^1 &= \frac{K}{\ln T} \int_1^T \frac{|G_t| |M_t - G_t|^\tau}{\sqrt{t} \sqrt{t}^\tau} \frac{dt}{t} \\ &\leq \frac{K}{\ln T} \int_1^T \frac{|G_t| |M_t| + |G_t| + 1}{\sqrt{t} \sqrt{t}^\tau} \frac{dt}{t} \text{ car } 0 \leq \tau \leq 1 \\ &\leq \frac{K}{\ln T} \int_1^T \frac{|G_t| |M_t|}{\sqrt{t} \sqrt{t}} \frac{dt}{t} + \frac{K}{\ln T} \int_1^T \frac{|G_t|^2}{\sqrt{t}^{\tau+1}} \frac{dt}{t} + \frac{K}{\ln T} \int_1^T \frac{|G_t|}{\sqrt{t}^{\tau+1}} \frac{dt}{t}. \end{aligned}$$

Le premier terme se traite de manière similaire à D_T^2 , les deux termes résiduels, comme D_T^3 . Par conséquent,

$$D_T^1 \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0$$

et notre preuve est complète. □

Remarquons que la fonction $x \mapsto x^2$ appartient à \mathcal{G}_1 . Dans ce cas, le principe d'invariance du théorème 3.2.1 prend la forme suivante :

COROLLAIRE 3.2.1. Sous les hypothèses (H1)–(H3),

$$(56) \quad \frac{1}{\ln T} \int_1^T \left(\frac{1}{t} \int_0^t \tilde{f}(X_s) ds \right)^2 dt \xrightarrow[T \rightarrow +\infty]{\mathbb{P}_x\text{-p.s.}} \sigma_f^2, \forall x \in \mathbb{R}^n.$$

Ce corollaire nous fournit un estimateur fortement consistant de la variance de notre simulation. Dans le cadre des chaînes de Markov récurrentes et possédant un atome, les auteurs de [12] proposent également un estimateur consistant de la variance ergodique. Cet estimateur reste toutefois très spécifique à ce cadre et ne semble pas pouvoir être adapté au notre.

3.3. Vitesse de convergence dans le principe d'invariance. La question naturelle est maintenant de donner une vitesse de convergence associée à la loi des grands nombres du théorème 3.2.1. Cette vitesse est donnée par le théorème suivant.

THÉORÈME 3.3.1. *Pour toute fonction φ appartenant à l'ensemble \mathcal{G}_τ défini au théorème 3.2.1 pour un $0 \leq \tau \leq 1$, nous posons*

$$\bar{\varphi} := \varphi - \int \varphi d\mathcal{N}(0, \sigma_f^2),$$

et, sous les hypothèses (H1)–(H3), nous avons alors

$$\frac{1}{\sqrt{\ln T}} \int_1^T \bar{\varphi} \left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \bar{\sigma}_\varphi^2)$$

avec

$$\bar{\sigma}_\varphi^2 := -2 \int \bar{\varphi} G_\varphi d\mathcal{N}(0, \sigma_f^2)$$

et G_φ , une solution du problème de Poisson

$$-y \dot{G}_\varphi(y) + \sigma_f^2 \ddot{G}_\varphi(y) = -2\bar{\varphi}.$$

Pour démontrer ce théorème, nous allons également nous inspirer de son équivalent pour les martingales locales. Ce résultat est établi par F. Maaouia [66] et F. Chaabane [22] sous une hypothèse supplémentaire. Le lemme suivant permet d'établir qu'elle est vérifiée dans notre cadre.

LEMME 3.3.1. *Avec les notations du théorème 3.2.1, sous les hypothèses (H1)–(H3),*

$$(\ln T)^2 \left(\frac{\langle M \rangle_T}{T} - \sigma_f^2 \right) \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0.$$

DÉMONSTRATION. Le théorème F de Maaouia [66] établit que notre lemme est une conséquence de l'hypothèse suivante sur la résolvante associée au semi-groupe du processus X :

(R-R) *Il existe une fonction positive f_0 , μ -intégrable, $\int f_0 d\mu > 0$, telle que, pour tout $x \in \mathbb{R}^n$, la fonction*

$$(57) \quad \lambda \mapsto \frac{1}{\lambda} \left(\lambda \int_0^{+\infty} e^{-\lambda t} \mathbb{E}_x [f_0(X_t)] dt - \int f_0 d\mu \right)$$

admette une limite finie lorsque λ tend vers 0.

Il est clair qu'il existe $f_0 > 0$ telle que $\int f_0 d\mu > 0$; il suffit par exemple de considérer l'indicatrice d'un borélien B tel que $\int_B d\mu > \frac{1}{2}$. Un tel ensemble existe toujours puisque μ est de masse totale un et son indicatrice vérifie bien l'hypothèse (H3). Par ailleurs, la fonction définie en (57) peut également s'écrire

$$\lambda \mapsto \int_0^{+\infty} e^{-\lambda t} \mathbb{E}_x [\tilde{f}_0(X_t)] dt.$$

Il est alors clair que, quand λ tend vers 0, cette fonction converge vers la solution $G_{f_0}(x)$ du problème de Poisson $\mathcal{L}G_{f_0} = -\tilde{f}_0$ dont nous savons qu'elle est finie pour tout f_0 vérifiant (H3) et tout $x \in \mathbb{R}^n$. D'où notre lemme. \square

PREUVE DU THÉORÈME 3.3.1. La preuve de ce théorème repose sur les mêmes idées que celle du théorème 3.2.1. Avec les mêmes notations, toujours d'après les résultats de F.

Maaouia [66] et F. Chaabane [22], nous savons que, sous l'hypothèse additionnelle

$$(58) \quad (\ln T)^2 \left(\frac{\langle M \rangle_T}{T} - \sigma_f^2 \right) \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0,$$

nous avons la vitesse de convergence du TCL presque sûr pour les martingales :

$$(59) \quad \frac{1}{\sqrt{\ln T}} \int_1^T \bar{\varphi} \left(\frac{M_t}{\sqrt{t}} \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \bar{\sigma}_\varphi^2).$$

Or, le lemme 3.3.1 nous assure que cette hypothèse est bien vérifiée dans notre cas. Par conséquent, sous les seules hypothèses (H1)–(H3), nous vérifions (59). Notre objectif est maintenant de montrer que la différence

$$D_T := \frac{1}{\sqrt{\ln T}} \int_1^T \left| \bar{\varphi} \left(\frac{M_t}{\sqrt{t}} \right) - \bar{\varphi} \left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right) \right| \frac{dt}{t}$$

converge presque sûrement vers 0. Au vu de la preuve du théorème 3.2.1, nous avons directement

$$D_T \leq \underbrace{\frac{K}{\sqrt{\ln T}} \int_1^T \frac{|G_t| |M_t - G_t|^\tau}{\sqrt{t} \sqrt{t}^\tau} \frac{dt}{t}}_{=: D_T^1} + \underbrace{\frac{K}{\sqrt{\ln T}} \int_1^T \frac{|G_t| |M_t|^\tau}{\sqrt{t} \sqrt{t}^\tau} \frac{dt}{t}}_{=: D_T^2} + \underbrace{\frac{K}{\sqrt{\ln T}} \int_1^T \frac{|G_t|}{\sqrt{t}} \frac{dt}{t}}_{=: D_T^3}$$

Et, toujours en suivant le raisonnement de la preuve précédente,

$$D_T^3 = \frac{1}{\sqrt{\ln T}} \frac{1}{T^{3/2}} \int_1^T |G_t| dt + \frac{3}{2\sqrt{\ln T}} \int_1^T \frac{1}{t^{3/2}} \int_1^t |G_s| ds \frac{dt}{t}.$$

Le premier terme ne pose aucun problème. Pour traiter le second terme, nous cherchons à faire apparaître une moyenne de Césaro. Comme la dérivée de $t \mapsto \sqrt{\ln t}$ est $t \mapsto \frac{1}{2t\sqrt{\ln t}}$, il est naturelle d'écrire

$$\frac{1}{\sqrt{\ln T}} \int_1^T \frac{1}{t^{3/2}} \int_1^t |G_s| ds \frac{dt}{t} = \frac{1}{\sqrt{\ln T}} \int_1^T \frac{\sqrt{\ln t}}{t^{3/2}} \int_1^t |G_s| ds \frac{dt}{t\sqrt{\ln t}}.$$

Comme nous l'avons déjà noté,

$$\frac{\sqrt{\ln t}}{t^{3/2}} \int_1^t |G_s| ds \xrightarrow[t \rightarrow +\infty]{\text{p.s.}} 0.$$

Par conséquent, nous avons bien

$$\frac{1}{\sqrt{\ln T}} \int_1^T \frac{\sqrt{\ln t}}{t^{3/2}} \int_1^t |G_s| ds \frac{dt}{t\sqrt{\ln t}} \xrightarrow[t \rightarrow +\infty]{\text{p.s.}} 0$$

et donc

$$D_T^3 \xrightarrow[t \rightarrow +\infty]{\text{p.s.}} 0.$$

Les termes D_T^1 et D_T^2 se traitent de manière identique. en faisant appel à l'équation (59). \square

Quand $\varphi(x) = x^2$ -c'est-à-dire lors de l'estimation de la variance ergodique- remarquons que $\bar{\sigma}_\varphi^2 = \sigma_f^2$. Nous obtenons alors le corrolaire établissant la normalité asymptotique de l'estimateur de la variance.

COROLLAIRE 3.3.1. Sous les hypothèses (H1)–(H3),

$$\frac{1}{\sqrt{\ln T}} \int_1^T \left(\left(\frac{1}{\sqrt{t}} \int_0^t \tilde{f}(X_s) ds \right)^2 - \sigma_f^2 \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_f^2).$$

La convergence de notre estimateur est logarithmique. Il est donc illusoire d’espérer une estimation précise de la variance en pratique. Néanmoins, cette procédure fournit pour un coût numérique négligeable (la simulation d’une unique trajectoire du processus de diffusion X est nécessaire), l’ordre de grandeur de la variance. A partir de cette donnée, il est donc possible de déterminer la durée d’une simulation nécessaire pour obtenir la précision voulue. Par ailleurs, comme la variance de notre estimateur est précisément la variance de la simulation, celle-ci sera d’autant mieux estimée qu’elle est petite.

3.4. Résultats numériques. Les figures 3, 4 et 5 illustrent la convergence de l'estimateur sur l'exemple du processus d'Ornstein-Uhlenbeck de paramètre θ avec la fonction $f(x) = x$.

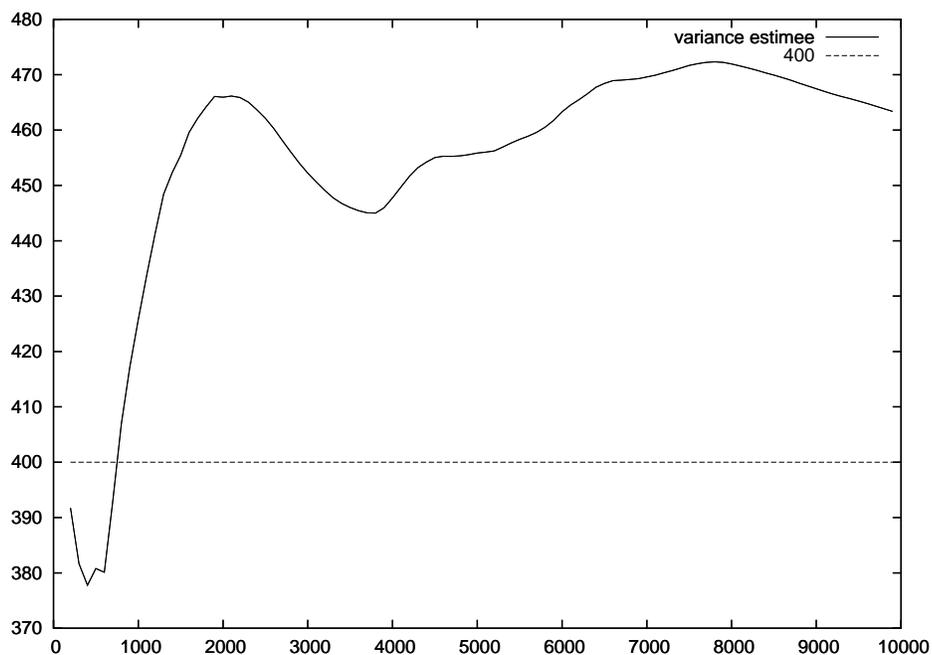


FIG. 3. Estimation de la variance pour $\theta = 0.05$

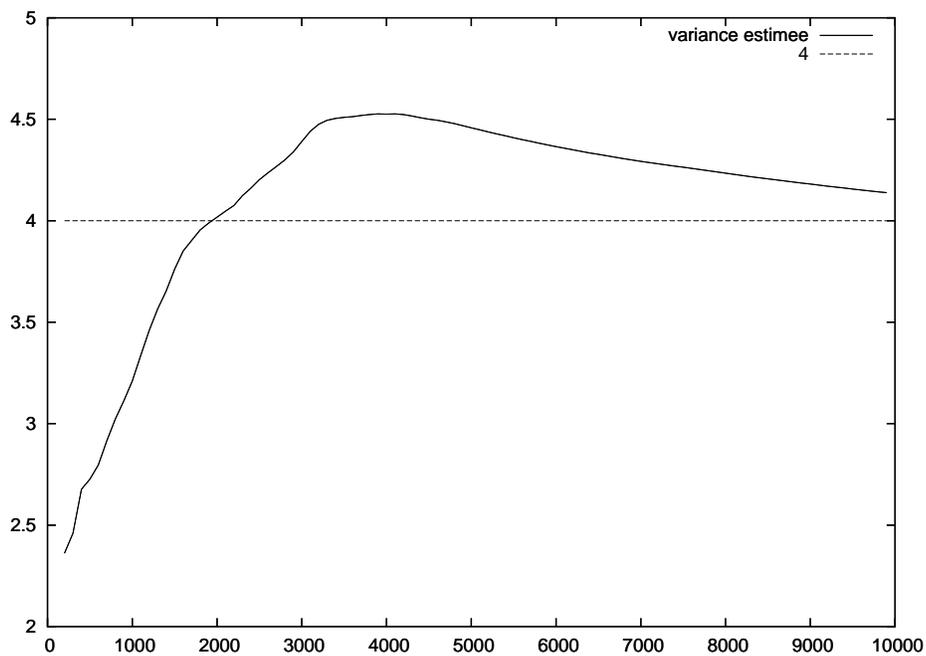


FIG. 4. Estimation de la variance pour $\theta = 0.5$

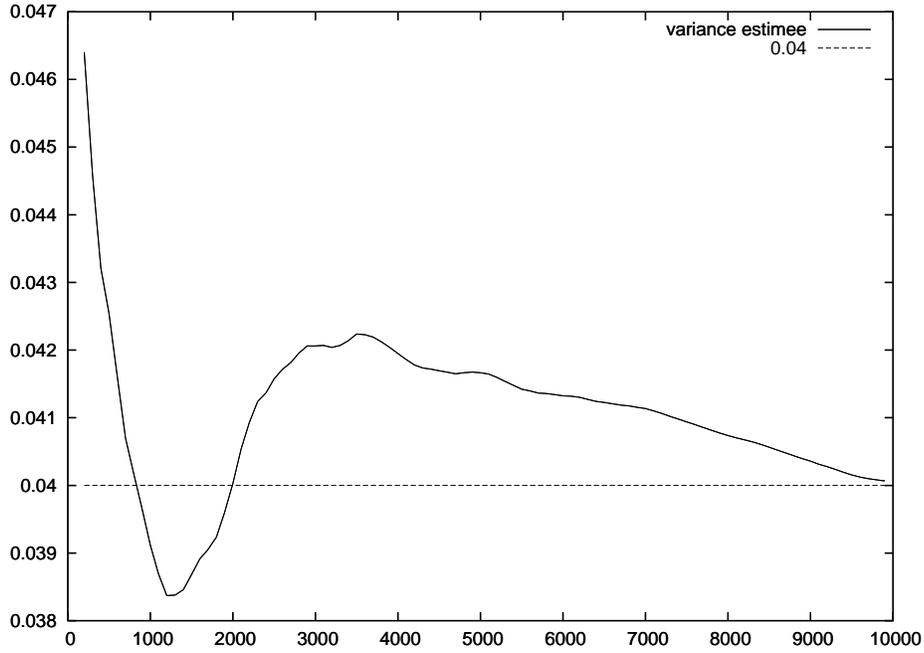


FIG. 5. Estimation de la variance pour $\theta = 5.0$

3.5. Une remarque : un théorème de la limite centrale presque sûr pour le temps local. Pour terminer cette section, nous énonçons un théorème de la limite centrale presque sûr pour le temps local L_t^x du processus X et donnons une vitesse de convergence associée.

THÉORÈME 3.5.1. *En dimension un, si l'hypothèse (43) est vérifiée, alors pour tout $0 \leq \tau \leq 1$ et toute fonction $\varphi \in \mathcal{G}_\tau$, avec \mathcal{G}_τ tel que défini au théorème 3.2.1,*

$$\frac{1}{\ln T} \int_1^T \varphi \left(\sqrt{t} \left(\frac{L_t^x}{t\sigma^2(x)} - p(x) \right) \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{p.s.} \int \varphi d\mathcal{N}(0, \sigma_x^2)$$

et

$$\frac{1}{\sqrt{\ln T}} \int_1^T \bar{\varphi} \left(\sqrt{t} \left(\frac{L_t^x}{t\sigma^2(x)} - p(x) \right) \right) \frac{dt}{t} \xrightarrow[T \rightarrow +\infty]{Loi} \mathcal{N}(0, \bar{\sigma}_\varphi^2)$$

avec σ_x^2 définie par (54),

$$\bar{\varphi} := \varphi - \int \varphi d\mathcal{N}(0, \sigma_x^2),$$

$$\bar{\sigma}_\varphi^2 := -2 \int \bar{\varphi} G_\varphi d\mathcal{N}(0, \sigma_x^2)$$

et G_φ , la solution du problème de Poisson

$$-y\dot{G}_\varphi(y) + \sigma_x^2 \ddot{G}_\varphi(y) = -2\bar{\varphi}.$$

DÉMONSTRATION. La démonstration est identique à celle des théorèmes 3.2.1 et 3.3.1, en prenant cette fois comme point de départ la représentation de Poisson (50). \square

4. Techniques de réduction de variance

Le but de cette dernière section est d'introduire trois techniques permettant de réduire la variance des simulations ergodiques. Pour illustrer chacune de ces techniques, nous considérerons un exemple très simple de processus stationnaire en dimension 2 :

$$(60) \quad dZ_t = B \cdot Z_t dt + \Sigma(Z_t) \cdot dW_t, \quad Z_0 \in \mathbb{R}^2, \quad t \geq 0$$

avec

$$B := \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}, \quad \alpha \in \mathbb{R}, \quad \beta \in \mathbb{R}, \quad \alpha^2 > \beta^2$$

et

$$\Sigma(z) := \begin{pmatrix} \sin(\nu(z_1 + z_2)) & \cos(\nu(z_1 + z_2)) \\ \sin(\nu(z_1 + z_2) + \frac{\pi}{3}) & \cos(\nu(z_1 + z_2) + \frac{\pi}{3}) \end{pmatrix}, \quad \nu \in \mathbb{R}.$$

Ce processus admet pour unique mesure invariante la loi gaussienne centrée de variance

$$\Gamma := \frac{1}{2(\alpha^2 - \beta^2)} \begin{pmatrix} \alpha - \frac{\beta}{2} & \frac{\alpha}{2} - \beta \\ \frac{\alpha}{2} - \beta & \alpha - \frac{\beta}{2} \end{pmatrix}.$$

4.1. Technique de corrélation.

4.1.1. *Principe.* Cette première technique est inspirée de la traditionnelle technique de variable de contrôle. Il s'agit en effet d'introduire un nouveau processus ergodique Y , de moyenne limite connue et que nous supposons nulle sans perte de généralité. Il est alors clair que

$$\frac{1}{T} \int_0^T f(X_s) ds - \frac{1}{T} \int_0^T Y_s ds \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \int f d\mu$$

où X désigne toujours le processus de diffusion solution de l'EDS (35). Pour que cette manipulation ait un intérêt, il est bien sûr nécessaire que les processus X et Y soient corrélés. Il est très simple de construire un processus Y répondant à ces impératifs, il suffit de choisir le processus d'Ornstein-Uhlenbeck solution de l'EDS

$$\begin{cases} Y_t &= -\vartheta \int_0^t Y_s ds + B_t, \quad t \geq 0, \quad \text{avec } \vartheta > 0 \\ B_t &= \sum_{i=1}^d \varrho_i W_t^i, \quad \text{avec } \sum_{i=1}^d \varrho_i^2 = 1. \end{cases}$$

Le problème est maintenant de déterminer les coefficients ϑ et $\varrho := (\varrho_i)_{i=1 \dots d}$ qui induisent une variance minimale. Pour ce faire, calculons la variance $\sigma_f^2(\vartheta, \varrho)$ de la simulation.

PROPOSITION 4.1.1. *Soit*

$$\sigma_f^2(\vartheta, \varrho) := \int_{\mathbb{R}^n} \left| (\sigma \nabla G_f)(x) - \frac{\varrho}{\vartheta} \right|^2 d\mu(x)$$

Sous les hypothèses (H1)–(H3), $\sigma_f^2(\vartheta, \varrho) < +\infty$ et

$$\frac{1}{\sqrt{T}} \int_0^T \left(\tilde{f}(X_s) - Y_s \right) ds \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N} \left(0, \sigma_f^2(\vartheta, \varrho) \right).$$

DÉMONSTRATION. La démonstration est inspirée de celle de la proposition 2.3.1. En utilisant la représentation (45) et l'équation vérifiée par Y , il est facile de voir que

$$\begin{aligned} \sqrt{T} \left(\frac{1}{T} \int_0^T f(X_s) ds - \frac{1}{T} \int_0^T Y_s ds - \int f d\mu \right) &= \frac{1}{\sqrt{T}} \left(\underbrace{\int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s - \frac{B_T}{\vartheta}}_{=: M_T} \right) \\ &+ \frac{1}{\sqrt{T}} \left(\underbrace{G_f(X_0) - G_f(X_T) + \frac{Y_T - Y_0}{\vartheta}}_{=: G_T} \right). \end{aligned}$$

Or, en utilisant les arguments de la preuve de la proposition 2.3.1, nous vérifions que

$$\frac{G_T}{\sqrt{T}} \xrightarrow[T \rightarrow +\infty]{\mathbb{P}} 0$$

et que

$$\frac{M_T}{\sqrt{T}} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_f^2(\vartheta, \varrho))$$

avec

$$\begin{aligned} \sigma_f^2(\vartheta, \varrho) &:= \text{p.s-} \lim_{T \rightarrow +\infty} \frac{\langle M \rangle_T}{T} \\ &= \text{p.s-} \lim_{T \rightarrow +\infty} \frac{1}{T} \left\langle \int_0^T (\sigma \nabla G_f)(X_s) \cdot dW_s - \frac{B_T}{\vartheta} \right\rangle_T \\ &= \text{p.s-} \lim_{T \rightarrow +\infty} \frac{1}{T} \left\langle \int_0^T \left((\sigma \nabla G_f)(X_s) - \frac{\varrho}{\vartheta} \right) \cdot dW_s \right\rangle_T \\ &= \text{p.s-} \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \left| (\sigma \nabla G_f)(X_s) - \frac{\varrho}{\vartheta} \right|^2 ds \\ &= \int_{\mathbb{R}^n} \left| (\sigma \nabla G_f)(x) - \frac{\varrho}{\vartheta} \right|^2 d\mu(x). \end{aligned}$$

Le lemme de Slutsky permet de conclure la preuve. \square

Par conséquent, pour obtenir la variance minimale, nous devons résoudre le problème de minimisation

$$\begin{cases} \min_{\vartheta, \varrho} \int_{\mathbb{R}^n} \left| (\sigma \nabla G_f)(x) - \frac{\varrho}{\vartheta} \right|^2 d\mu(x) \\ \text{sous les contraintes } \vartheta > 0 \text{ et } \sum_{i=1}^d \varrho_i^2 = 1. \end{cases}$$

C'est un problème de minimisation convexe dont une solution est donnée par la

PROPOSITION 4.1.2.

$$\begin{aligned} \bar{\vartheta} &= \frac{1}{\sqrt{\sum_{i=1}^n \left(\int (\sigma \nabla G_f)_i d\mu \right)^2}} \\ \bar{\varrho}_i &= \frac{\int (\sigma \nabla G_f)_i d\mu}{\sqrt{\sum_{i=1}^n \left(\int (\sigma \nabla G_f)_i d\mu \right)^2}}, \quad i = 1 \dots n \end{aligned}$$

DÉMONSTRATION. Nous cherchons à minimiser

$$\int \left| (\sigma \nabla G_f) - \frac{\varrho}{\vartheta} \right|_2^2 d\mu = \int |(\sigma \nabla G_f)|_2^2 d\mu - \frac{2\varrho}{\vartheta} \int (\sigma \nabla G_f) d\mu + \left| \frac{\varrho}{\vartheta} \right|_2^2.$$

Il est clair que le minimum unique de cette équation du second ordre est atteint pour

$$\frac{\varrho}{\vartheta} = \int (\sigma \nabla G_f) d\mu.$$

Il nous est alors possible de choisir ϱ et ϑ pour vérifier les contraintes du problème, d'où les expressions

$$\bar{\varrho} = \frac{\int (\sigma \nabla G_f) d\mu}{\left| \int (\sigma \nabla G_f) d\mu \right|_2} \text{ et } \bar{\vartheta} = \frac{1}{\left| \int (\sigma \nabla G_f) d\mu \right|_2}.$$

□

La variance optimale est donc

$$\begin{aligned} \sigma_f^2(\bar{\vartheta}, \bar{\varrho}) &= \int |\sigma \nabla G_f|_2^2 d\mu - \left| \int \sigma \nabla G_f d\mu \right|_2^2 \\ &= \text{Var}_\mu [\sigma \nabla G_f]. \end{aligned}$$

Cette variance n'est donc pas nulle en général. Par ailleurs, la mise en œuvre de l'algorithme requiert la connaissance a priori de la moyenne $\int \sigma \nabla G_f d\mu$. Notons enfin que l'erreur de discrétisation n'est pas affectée par cette transformation.

4.1.2. *Domaine d'application de la méthode et illustration.* Du fait que la variance minimale soit $\text{Var}_\mu [\sigma \nabla G_f]$, il est clair que cette méthode fournira de bons résultats pour les cas où la fonction $\sigma \nabla G_f$ est proche d'une constante. Ce sera par exemple le cas pour une fonction de diffusion σ proche d'une constante (ce qui n'est pas trop demander au vu de nos hypothèses) et une dérive b dont la somme des composantes est proportionnelle à \tilde{f} . Considérons l'exemple du processus Z défini par l'EDS (60) et de la fonction $f(z) = z_1 + z_2$. Nous avons alors $\int f d\mu = 0$. Les figures 6, 7, 8 et 9 illustrent l'efficacité de notre méthode sur cet exemple. Le pas de discrétisation est $h = 10^{-3}$.

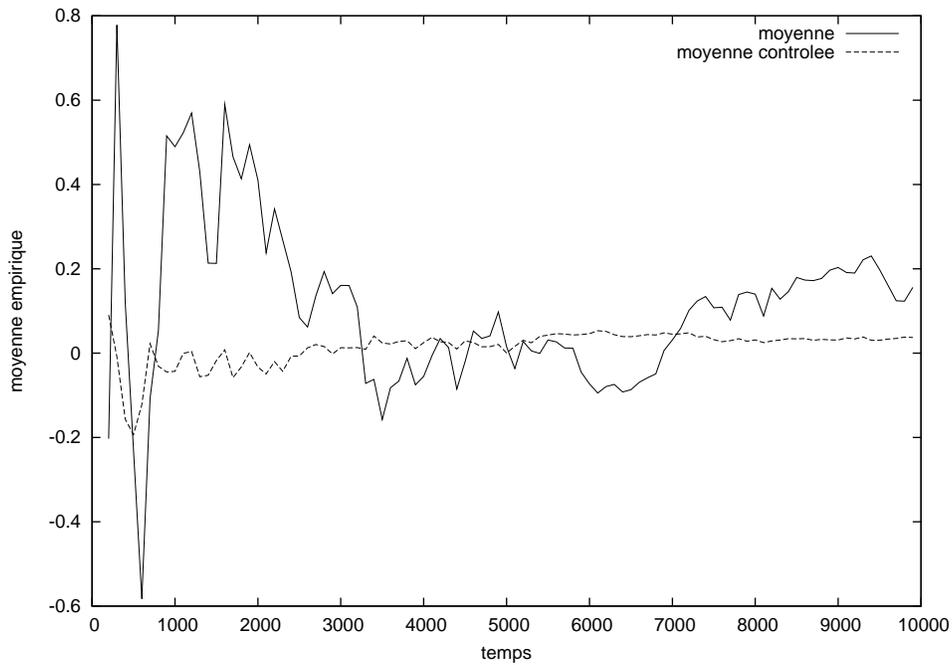


FIG. 6. Moyennes estimées pour $\alpha = -0.05$, $\beta = -0.025$, $\nu = 0.0$, $\vartheta = 0.043$, $\varrho_1 = 0.5$ et $\varrho_2 = 0.87$.

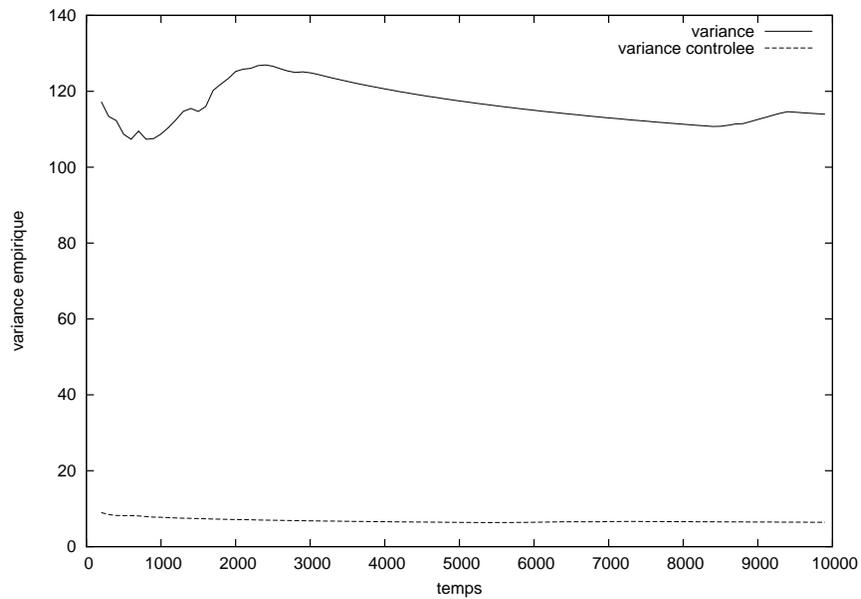


FIG. 7. Variances estimées pour $\alpha = -0.05$, $\beta = -0.025$, $\nu = 0.0$, $\vartheta = 0.043$, $\varrho_1 = 0.5$ et $\varrho_2 = 0.87$.

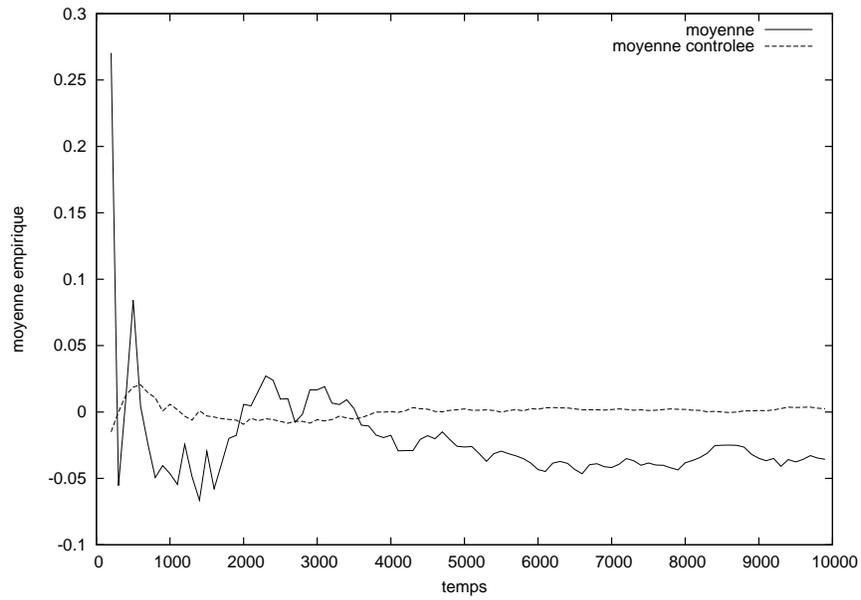


FIG. 8. Moyennes estimées pour $\alpha = -0.5$, $\beta = -0.25$, $\nu = 0.2$, $\vartheta = 0.43$, $\varrho_1 = 0.5$ et $\varrho_2 = 0.87$.

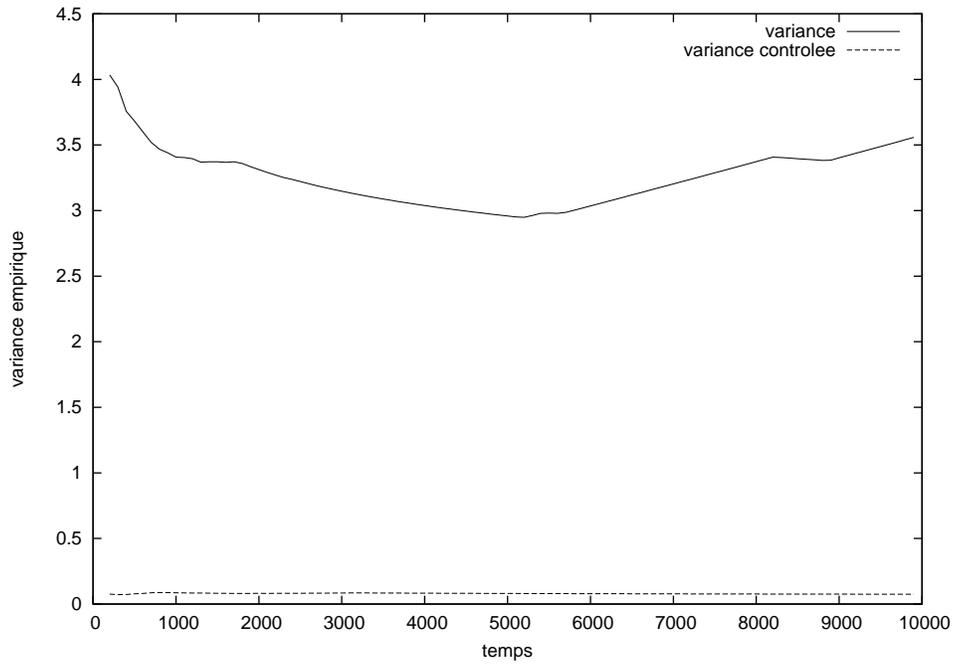


FIG. 9. Variances estimées pour $\alpha = -0.5$, $\beta = -0.25$, $\nu = 0.2$, $\vartheta = 0.43$, $\varrho_1 = 0.5$ et $\varrho_2 = 0.87$.

4.2. Technique d'accélération. Nous proposons maintenant une méthode plus spécifique aux simulations ergodiques; nous n'en connaissons en effet aucun équivalent pour les simulations Monte-Carlo classiques. Cette technique repose sur un changement de temps. Introduisons donc une fonction $\kappa : \mathbb{R}^n \rightarrow \mathbb{R}$ de $L^2(\mu)$, strictement positive. Nous définissons le changement de temps

$$h(t) := \int_0^t \kappa^2(X_s) ds, \quad t \geq 0,$$

et le processus $\tilde{X}_t := X_{h^{-1}(t)}$. Le processus \tilde{X} est alors solution de l'EDS suivante (voir Ikeda et Watanabe [44]) :

$$d\tilde{X}_t = \kappa(\tilde{X}_t)^{-2} b(\tilde{X}_t) dt + \kappa(\tilde{X}_t)^{-1} \sigma(\tilde{X}_t) d\tilde{W}_t, \quad \tilde{X}_0 = X_0, \quad t \geq 0$$

où $\tilde{W}_t := \int_0^{h^{-1}(t)} \kappa(X_s) dW_s$ est un nouveau mouvement brownien standard sur \mathbb{R}^d . Ce processus \tilde{X} admet alors une unique mesure invariante (voir toujours Ikeda et Watanabe [44]) dont la densité est $\tilde{p} \propto \kappa^2 p$. Il est alors possible d'introduire un nouvel estimateur de $\int f d\mu$.

PROPOSITION 4.2.1. *Sous les hypothèses (H1)–(H3),*

$$\frac{\int_0^T (\kappa^{-2} f)(\tilde{X}_s) ds}{\int_0^T \kappa^{-2}(\tilde{X}_s) ds} \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \int f d\mu.$$

DÉMONSTRATION. La preuve est immédiate car, d'après le théorème ergodique, d'une part

$$\begin{aligned} \frac{1}{T} \int_0^T (\kappa^{-2} f)(\tilde{X}_s) ds &\xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \int_{\mathbb{R}^n} \kappa^{-2}(x) f(x) \tilde{p}(x) dx \\ &= c \int_{\mathbb{R}^n} \kappa^{-2}(x) f(x) \kappa^2(x) p(x) dx \text{ avec } c := \left(\int \kappa^2 d\mu \right)^{-1} \\ &= c \int f d\mu, \end{aligned}$$

et d'autre part, pour les mêmes raisons,

$$\frac{1}{T} \int_0^T \kappa^{-2}(\tilde{X}_s) ds \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} c.$$

□

4.2.1. *Erreur Statistique.* La variance de la simulation est donnée par la proposition suivante.

PROPOSITION 4.2.2. *Sous les hypothèses (H1)–(H3),*

$$\sqrt{T} \frac{\int_0^T \kappa^{-2} \tilde{f}(\tilde{X}_s) ds}{\int_0^T \kappa^{-2}(\tilde{X}_s) ds} \xrightarrow[T \rightarrow \infty]{\text{Loi}} \mathcal{N} \left(0, \sigma_f^2 \int \kappa^2 d\mu \right).$$

DÉMONSTRATION. Il suffit encore une fois de vérifier à l'aide du théorème ergodique 1.0.1 que

$$\frac{1}{T} \int_0^T \kappa^{-2}(\tilde{X}_s) ds \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \left(\int \kappa^2 d\mu \right)^{-1}$$

et à l'aide du TCL 2.3.1 que

$$\frac{1}{\sqrt{T}} \int_0^T \kappa^{-2} \tilde{f}(\tilde{X}_s) ds \xrightarrow[T \rightarrow \infty]{\text{Loi}} \mathcal{N} \left(0, \frac{\int |\sigma \nabla G_f|^2 d\mu}{\int \kappa^2 d\mu} \right).$$

Le lemme de Slutsky conclut la preuve. \square

Pour optimiser notre simulation, il nous faut donc minimiser $\int \kappa^2 d\mu$. Sur ce seul critère, la logique suggérerait de choisir une fonction κ aussi proche de 0 que possible. Comme nous allons le voir sur l'étude de l'erreur de discrétisation, ce choix n'est pas judicieux.

4.2.2. *Erreur de discrétisation.* Pour mettre en évidence l'influence du changement de temps sur l'erreur de discrétisation, nous allons nous restreindre à un cas très simple. Posons $\kappa(t) := \frac{1}{\lambda}$ avec $\lambda > 0$. Le processus \tilde{X} vérifie alors l'EDS suivante :

$$\tilde{X}_t = X_0 + \lambda^2 \int_0^t b(\tilde{X}_s) ds + \lambda \int_0^t \sigma(\tilde{X}_s) d\tilde{W}_s.$$

Il est facile de vérifier que les processus X et \tilde{X} ont même loi invariante et que, d'une part

$$\frac{1}{T} \int_0^T f(\tilde{X}_s) ds \xrightarrow[T \rightarrow \infty]{\text{p.s.}} \int f d\mu,$$

et d'autre part

$$\frac{1}{\sqrt{T}} \int_0^T \tilde{f}(\tilde{X}_s) ds \xrightarrow[T \rightarrow \infty]{\text{Loi}} \mathcal{N} \left(0, \frac{1}{\lambda^2} \int (\sigma \nabla G_f)^2 d\mu \right).$$

Par conséquent, prendre λ arbitrairement grand permettrait en théorie d'améliorer la simulation. Considérons maintenant le schéma de discrétisation du processus initial X :

$$\begin{cases} \bar{X}_0^h &= X_0 \\ \bar{X}_{p+1}^h &= \bar{X}_p^h + b(\bar{X}_p^h)h + \sigma(\bar{X}_p^h)W_p^h, p > 0, \end{cases}$$

où les W_p^h sont des gaussiennes indépendantes, centrées et de variance h . Il est connu (voir D. Talay et L. Tubaro [85]) que – sous des hypothèses de régularité supplémentaires sur les coefficients b et σ , et sur la fonction f – ce schéma converge à la vitesse h au sens où

$$\frac{1}{N} \sum_{p=1}^N f(\bar{X}_p^h) \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \int f d\mu + e_1 h + O(h^2).$$

Ce résultat appliqué au schéma d'Euler correspondant au processus \tilde{X} devient

$$\frac{1}{N} \sum_{p=1}^N f(\tilde{\bar{X}}_p^h) \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \int f d\mu + e_1 \lambda^2 h + O(h^2).$$

Le changement de temps apparaît clairement ici. Tout comme il apparaît clairement que choisir un paramètre λ trop grand compromet la précision de la simulation. En pratique, nous avons donc à résoudre un dilemme biais-variance. Pour tenir compte du biais, il est

possible de considérer un critère reposant sur l'erreur quadratique et non plus sur la seule variance. Il s'agit alors de déterminer la valeur de λ qui minimise

$$e_1^2 \lambda^4 h^2 + \frac{1}{\lambda^2} \int |\sigma \nabla G_f|^2 d\mu.$$

Cette valeur optimale est

$$\bar{\lambda} := \left(\frac{\int |\sigma \nabla G_f|^2 d\mu}{2e_1^2 h^2} \right)^{\frac{1}{6}}.$$

4.2.3. *Illustration.* Considérons à nouveau le processus Z défini en (60) et la fonction $f(z) = z_1^2 + z_2^2$. Les figures 10 et 11 illustrent la technique d'accélération précédemment décrite. Le pas de discrétisation est $h = 10^{-3}$. La valeur exacte est $\int f d\mu = 0.2$.

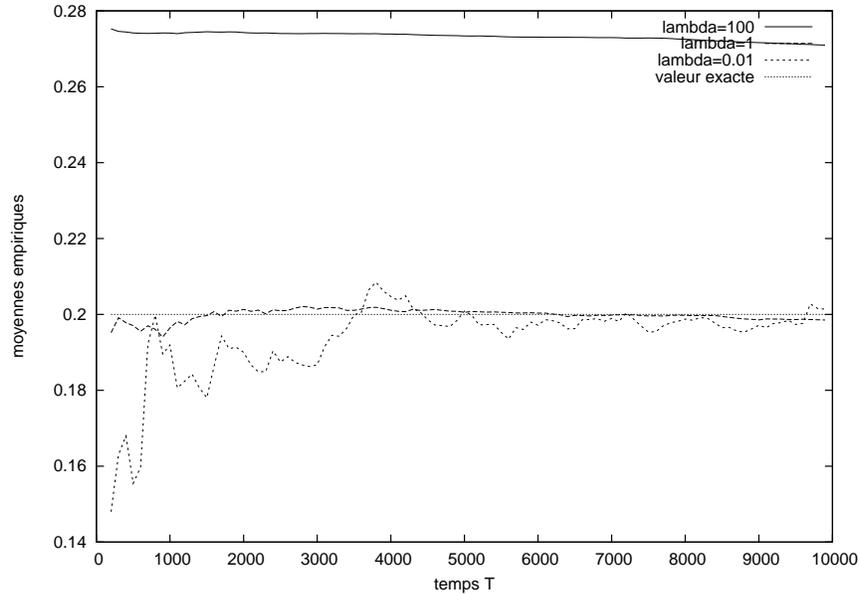


FIG. 10. Moyennes estimées pour $\alpha = -5.0$, $\beta = -2.5$, $\nu = 0.2$ et $\lambda \in \{0.01, 1.0, 100.0\}$.

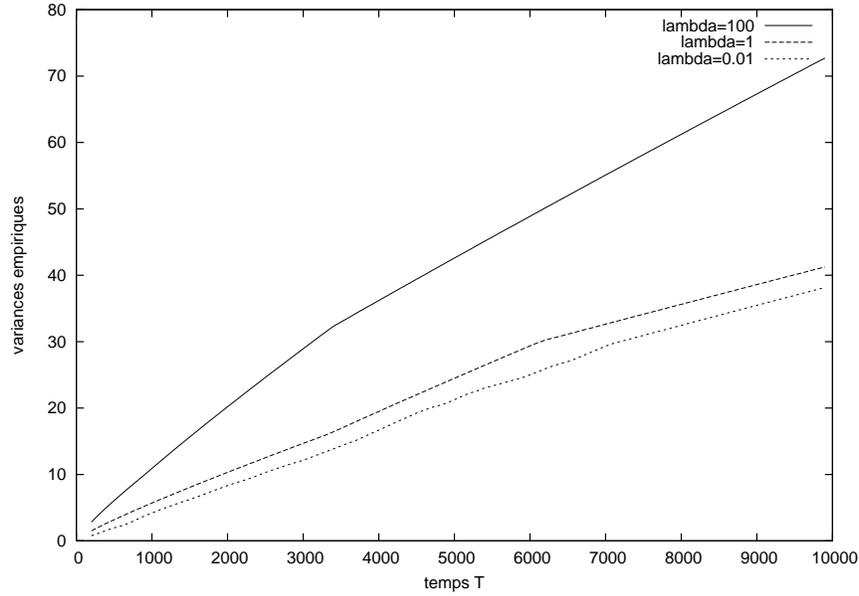


FIG. 11. Variances estimées pour $\alpha = -5.0$, $\beta = -2.5$, $\nu = 0.2$ et $\lambda \in \{0.01, 1.0, 100.0\}$.

4.3. Technique de régularisation. Il s'agit ici d'adapter une méthode de réduction de variance utilisée depuis peu pour les simulations Monte-Carlo, en particulier dans le domaine de la finance (voir entre autres [34],[33], [39], [48] et [49]).

4.3.1. *Intégration par parties.* Quand la fonctionnelle dont nous recherchons l'espérance est peu régulière, il est possible d'avoir recours à la transformation suivante.

PROPOSITION 4.3.1. *Si la mesure μ admet une densité p de classe C^2 alors, pour toutes fonctions $f \in C^1(\mathbb{R}^n, \mathbb{R})$ et $g \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ telles que,*

$$|fg\nabla p|(x) \xrightarrow{|x| \rightarrow +\infty} 0 \text{ et } |f \operatorname{div} g p|(x) \xrightarrow{|x| \rightarrow +\infty} 0$$

nous avons, presque sûrement,

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \nabla f(X_s) \cdot g(X_s) ds$$

=

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(X_s) (-\operatorname{div} g(X_s) + 2g(X_s) \cdot a^{-1}\bar{b}) ds$$

avec $\bar{b}_i := \frac{1}{2} \sum_{j=1}^n \frac{\partial a_{ij}}{\partial x_j} - b_i$, $i = 1 \dots n$.

Notons dès à présent que l'utilisation de cette méthode requiert l'inversion de la matrice a et l'évaluation du vecteur \bar{b} . Elle impose donc potentiellement un surcoût numérique non négligeable. Une localisation est possible par l'intermédiaire de la fonction g .

DÉMONSTRATION. L'équation de Fokker-Planck (36) s'écrit sous forme condensée

$$\operatorname{div} \left[\frac{1}{2} a \dot{p} + \bar{b} p \right] = 0.$$

Nous savons que la densité p est l'unique solution de cette équation. Il est alors facile de vérifier, du fait de l'inversibilité de la matrice a , que p doit vérifier

$$\frac{\dot{p}}{p} = -2a^{-1}\bar{b}.$$

La proposition est alors une simple conséquence de la formule d'intégration par parties

$$\int_{\mathbb{R}^n} \nabla f(y) \cdot g(y)p(y) dy = \int_{\mathbb{R}^n} (-f(y) \operatorname{div} g(y) p(y) + f(y)g(y) \cdot \nabla p(y)) dy$$

et du théorème ergodique. □

4.3.2. *Illustration.* Cette dernière méthode est plus difficile à mettre en œuvre et plus coûteuse numériquement que les deux précédentes. Nous nous contenterons de l'illustrer sur un cas très simple, l'estimation de la première marginale du processus Z défini en (60). Nous allons en fait comparer la méthode d'intégration par parties à la méthode traditionnelle de noyau. Rappelons que Castellana et Leadbetter [21] (voir aussi Bosq [15]) ont établi l'existence d'estimateurs non-paramétriques de la densité convergeant avec un taux paramétrique \sqrt{T} . Nous nous autoriserons donc à ne considérer que la variance de la simulation comme critère de comparaison. Les paramètres de la simulation sont les suivants :

$$Z_0 = 0, \alpha = \frac{1}{2}, \beta = 0, \nu = 0, h = 10^{-4}.$$

Ainsi, la densité que nous cherchons à estimer est celle d'une loi normale centrée réduite. Le choix des fonctions de localisation est une question délicate. Pour l'estimation par noyau, il est connu que le noyau d'Epanechnikov est optimal au sens de l'erreur quadratique, c'est donc celui que nous avons retenu. Nous ne sommes pas parvenu à établir de résultat théorique concernant la localisation optimale pour la méthode d'intégration par parties. Nous avons donc choisi arbitrairement une fonction de localisation gaussienne. En pratique, plus que le noyau, c'est la taille de la fenêtre qui importe. Dans les deux cas, la taille optimale a été déterminée par validation croisée.

La figure 12 représente les densités estimées par la méthode de noyau et par intégration par parties, avec et sans localisation ($g \equiv 1$ dans ce dernier cas). La figure 13 illustre la convergence de l'erreur quadratique en fonction de la durée de simulation pour chacune de ces trois méthodes. Enfin, la figure 14 représente la variance estimée pour la méthode de noyau et la méthode d'intégration par parties avec localisation. Cette dernière figure est à comparer avec le résultat théorique 1 de la page 27.

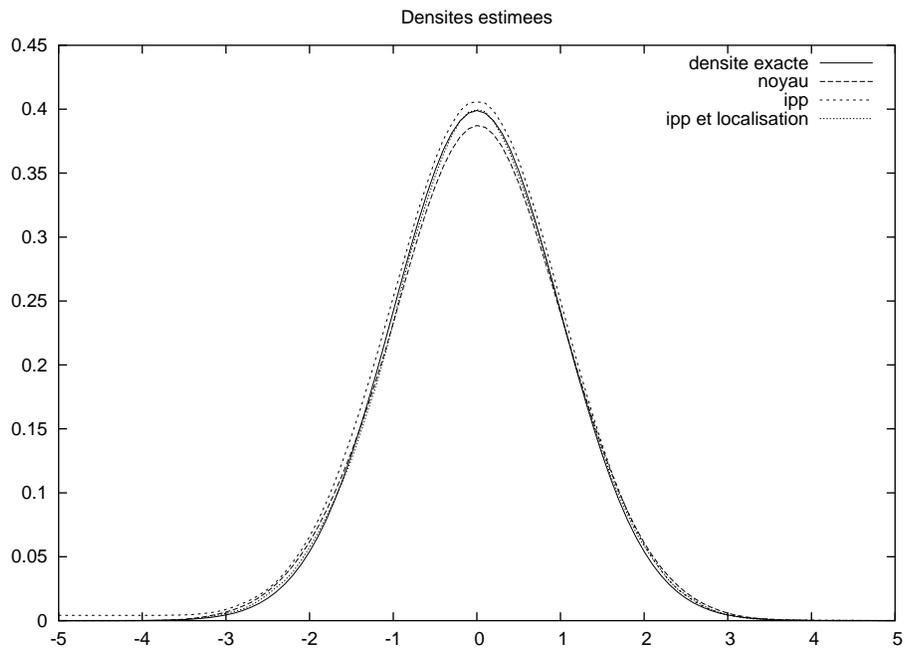


FIG. 12. Densités estimées

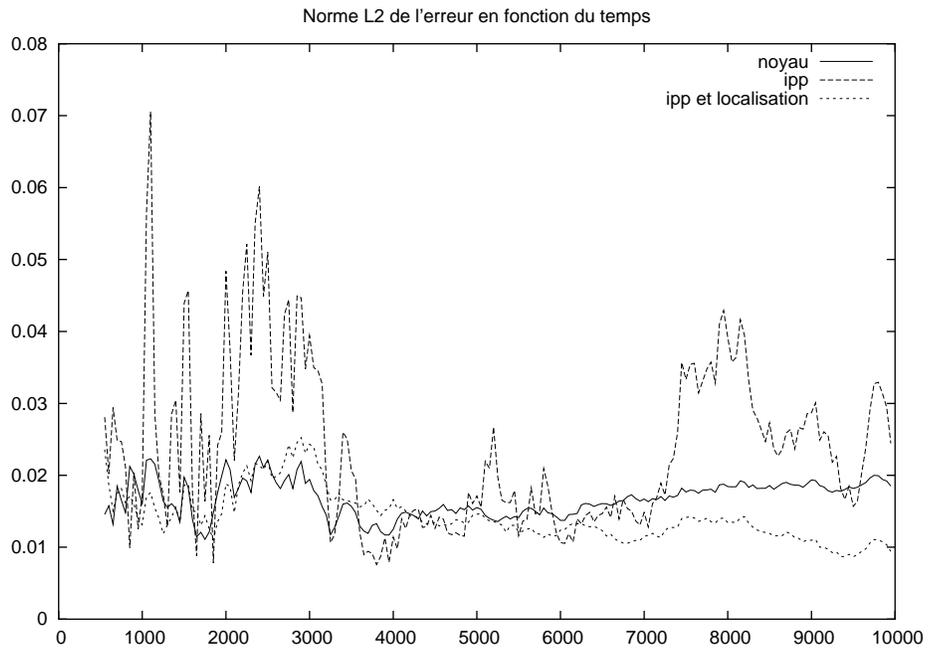


FIG. 13. Evolution de l'erreur quadratique en fonction du temps

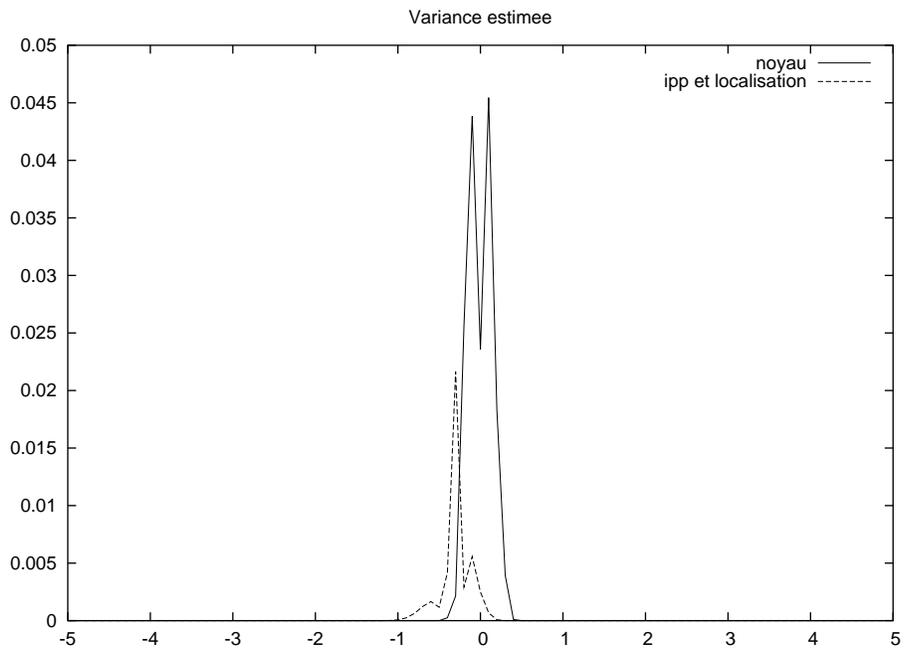


FIG. 14. Variances estimées

Il semble que, sur cet exemple particulier, la méthode d'intégration par parties donne les meilleurs résultats.

5. Conclusion et perspectives

Dans ce chapitre, nous nous sommes efforcés d'introduire certaines idées permettant de traiter les algorithmes de simulation ergodique du point de vue de la réduction de variance. Nous avons ainsi présenté un estimateur simple de la variance asymptotique, quantité qui ne pouvait jusqu'à présent être estimée que par des procédures complexes et coûteuses. Par ailleurs, trois techniques de réduction de variance ont été construites dans le cadre ergodique. A notre connaissance, il n'existe pas d'approche similaire. Beaucoup reste donc à faire. Il nous semble que la question la plus difficile est l'optimisation des différentes techniques de réduction de variance. C'est un problème extrêmement complexe, même dans le cadre des méthodes de Monte-Carlo classiques. Pour l'aborder efficacement dans le contexte ergodique, il semble nécessaire d'élaborer de nouveaux outils permettant d'obtenir des informations a priori sur la variance de la simulation.

Deuxième partie

Deux contributions à la Statistique des
diffusions

Cette deuxième partie consiste en deux chapitres.

- Au cours du premier chapitre, nous établissons des développements d'Edgeworth pour trois familles de fonctionnelles de diffusions ergodiques : les martingales, les martingales normalisées et les moyennes empiriques. Nous présentons en outre différentes applications de ces résultats en statistique des processus et en simulation.
- Le second chapitre aborde le problème de l'estimation paramétrique des processus de diffusions asymétriques généralisant le mouvement brownien asymétrique. Notre approche repose sur une propriété d'ergodicité du modèle. En combinant une méthode de moments et une procédure de minimisation de contraste, nous montrons qu'il est possible d'estimer de manière adaptative et consistante à la fois le lieu et l'amplitude de l'asymétrie.

DÉVELOPPEMENTS D'EDGEWORTH POUR LES FONCTIONNELLES DE PROCESSUS DE DIFFUSION ERGODIQUES

1. Introduction

Au cours de ce chapitre, nous nous proposons de préciser différents résultats de convergence de fonctionnelles de processus de diffusion en établissant des développements d'Edgeworth de leur fonction de répartition. Diverses applications en simulation et en statistique des processus seront étudiées.

1.1. Hypothèses. Nous considérons le processus X de \mathbb{R}^n , solution de

$$(61) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = x$$

où W est un mouvement brownien standard sur \mathbb{R}^d . Nous appellerons \mathcal{L} le générateur infinitésimal du processus X . Nous nous plaçons sous les hypothèses suivantes où le \cdot désigne le produit scalaire dans \mathbb{R}^n et $|\cdot|$, la norme euclidienne dans \mathbb{R}^n ,

(H1) La fonction a est bornée, uniformément continue et uniformément elliptique :

$$\exists K > 0 / \forall x, \xi \in \mathbb{R}^n, \sum_{i,j} a_{i,j}(x) \xi_i \xi_j \geq K |\xi|^2.$$

(H2) La fonction b est localement bornée et il existe $\alpha > -1$, $M > 0$ et $r > 0$ tels que

$$b(x) \cdot \frac{x}{|x|} \leq -r|x|^\alpha, \quad |x| \geq M.$$

(H3) La fonction $g : \mathbb{R}^n \rightarrow \mathbb{R}^d$ est continue, n'est pas constante et il existe $\beta \geq 0$, $C_1 > 0$ et $C_2 > 0$ tels que, pour tout $x \in \mathbb{R}^n$,

$$|g(x)| \leq C_1 + C_2|x|^\beta.$$

(H4) Pour toute fonction g vérifiant l'hypothèse (H3), il existe des réels $T_0 > 0$, $\omega_0 > 0$ et $\gamma > 0$ tels que, pour tout $|\omega| > \omega_0$,

$$|\mathbb{E} [\exp(i\omega M_T^g)]| \leq e^{-\gamma T}, \quad \forall T \geq T_0$$

où M^g est la martingale locale définie par

$$M_T^g := \int_0^T g(X_s) \cdot dW_s, \quad T \geq 0.$$

1.2. Premières conséquences des hypothèses (H1)–(H3). Ce sont les hypothèses de travail de E. Pardoux et A. Yu. Veretennikov [72], le cas $\alpha = -1$ exclu. Ce dernier cas est plus techniques car il impose des conditions supplémentaires entre les différentes constantes introduites dans les hypothèses, en particulier pour assurer l'existence des moments du processus X . Nous écartons ce cas pour ne pas surcharger la présentation mais tous nos résultats s'étendent sans difficulté à cette hypothèse.

Nous commençons par rappeler certains des résultats de [72] que nous utiliserons continuellement par la suite.

PROPOSITION 1.2.1. *Sous les hypothèses (H1)–(H3), pour tout $m > 0$, il existe des constantes $C_1 > 0$, $C_2 > 0$ et $C_3 > 0$ indépendantes du temps et telles que*

$$(i) \quad \mathbb{E}_\mu [|X_t|^m] = C_1 < +\infty,$$

$$(ii) \quad \mathbb{E}_x [|X_t|^m] = C_2(1 + |x|^m)$$

et pour tout $2k + 2 < m$, $k > 0$,

$$(iii) \quad \text{var}(\mu_t^x - \mu) \leq \frac{C_3(1 + |x|^m)}{(1 + t)^{k+1}}$$

où var désigne la norme en variation quadratique d'une mesure signée sur la sigma algèbre, μ_t^x est la loi de X_t quand $X_0 = x$ et μ est l'unique mesure invariante de X . De plus, quand $\alpha > 0$, il existe des constantes $C > 0$, $\nu > 0$ et $\lambda > 0$ telles que, pour tout $t \geq 0$,

$$(iv) \quad \mathbb{E}_x [\exp(\nu|X_t|)] \leq C \exp(\nu|x|),$$

$$(v) \quad \mathbb{E}_\mu [\exp(\nu|X_t|)] < +\infty$$

et

$$(vi) \quad \text{var}(\mu_t^x - \mu) \leq C \exp(\nu|x|) \exp(-\lambda t).$$

DÉMONSTRATION. Voir [72] et les références citées. □

En particulier, toute fonction f vérifiant (H3) est μ -intégrable et nous poserons

$$\mu(f) := \int f d\mu \text{ et } \tilde{f} := f - \mu(f).$$

En guise de préliminaire, nous établissons la proposition suivante qui caractérise la vitesse de convergence du semi-groupe vers la mesure invariante.

PROPOSITION 1.2.2. *Nous nous plaçons sous les hypothèses (H1)–(H3) et nous définissons*

$$p_f(t, x) := \mathbb{E}_x [\tilde{f}(X_t)].$$

Pour tout $k > 0$, il existe des constantes strictement positives C_0 et s_0 telles que

$$|p_f(t, x)| \leq \frac{C_0(1 + |x|^{s_0})}{(1 + t)^{k+1}}.$$

DÉMONSTRATION. Pour prouver cette proposition, nous reprenons les éléments de la preuve du théorème 1 de [72]. Les constantes C_0 , C_1 et C_2 peuvent varier d'une ligne à

l'autre,

$$\begin{aligned} |p_f(t, x)| &= \left| \int \tilde{f}(y) \mu_t^x(dy) \right| \\ &= \left| \int \tilde{f}(y) (\mu_t^x(dy) - \mu(dy)) \right|. \end{aligned}$$

Nous appliquons maintenant l'inégalité de Hölder avec $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned} |p_f(t, x)| &\leq \left(\int |\tilde{f}(y)|^p (\mu_t^x(dy) + \mu(dy)) \right)^{\frac{1}{p}} \left(\int |\mu_t^x - \mu|(dy) \right)^{\frac{1}{q}} \\ &\leq C_0 \left(\int (C_1 + C_2|y|^m) (\mu_t^x(dy) + \mu(dy)) \right)^{\frac{1}{p}} \text{var}(\mu_t^x - \mu)^{\frac{1}{q}} \end{aligned}$$

en utilisant l'hypothèse (H3) avec $m := \beta p$. En faisant maintenant appel à la proposition 1.2.1, nous obtenons, pour $h + 1 < \frac{m}{2}$,

$$\begin{aligned} |p_f(t, x)| &\leq C_0 (2C_1 + C_2 \mathbb{E}_x[|X_t|^m] + C_2 \mathbb{E}_\mu[|X_t|^m])^{\frac{1}{p}} ((1 + |x|^m)(1 + t)^{-(h+1)})^{\frac{1}{q}} \\ &\leq C_0 (1 + |x|^m)^{\frac{1}{q}} (C_1 + C_2(1 + |x|^m))^{\frac{1}{p}} (1 + t)^{-\frac{h+1}{q}} \\ &\leq C_0 (1 + |x|^m) (1 + t)^{-(k+1)} \end{aligned}$$

après avoir posé $k := \frac{h+1}{q} - 1$. Comme p est arbitraire, h l'est également et nous vérifions bien le résultat annoncé. \square

1.3. Quelques résultats de convergence pour les fonctionnelles de diffusion.

Nous établissons maintenant quelques conséquences de ces deux propositions. Rappelons que pour toute fonction $g : \mathbb{R}^n \rightarrow \mathbb{R}^d$ vérifiant (H3), nous avons défini

$$M_T^g := \int_0^T g(X_s) \cdot dW_s$$

et par conséquent

$$\langle M^g \rangle_T = \int_0^T |g(X_s)|^2 ds.$$

Notons dès à présent que les résultats de la proposition 1.2.1 nous assurent que les moments de tous ordres des martingales M^g sont finis.

1.3.1. *Convergence des moyennes empiriques.* D'après la proposition 1.2.1 le processus X est ergodique et, pour toute fonction f vérifiant (H3), nous avons le théorème ergodique (voir Has'minskiï [42] et Kutoyants [55])

$$(62) \quad \frac{1}{T} \int_0^T f(X_s) ds \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \mu(f).$$

Nous avons vu au cours du chapitre 2 qu'il était possible d'obtenir une vitesse de convergence associée en utilisant la solution de l'équation de Poisson

$$(63) \quad \mathcal{L}G_f = -\tilde{f}.$$

Les résultats de Pardoux et Veretennikov [72] nous assurent qu'il existe une unique fonction $G_f \in \bigcup_{p>1} W_{p,loc}^2$, à croissance polynomiale et vérifiant $\int G_f d\mu = 0$, solution de l'équation (63). Cette solution a pour expression

$$(64) \quad G_f(x) = \int_0^{+\infty} p_f(t, x) dt$$

avec les notations de la proposition 1.2.2. Une conséquence immédiate de la formule d'Itô-Krylov est la représentation

$$(65) \quad \int_0^T f(X_s) ds - \mu(f)T = M_T^{\sigma^* \nabla G_f} + G_f(x) - G_f(X_T).$$

Il est alors immédiat d'établir (voir le chapitre 2) le théorème de la limite centrale

$$(66) \quad \frac{1}{\sqrt{T}} \int_0^T \tilde{f}(X_s) ds \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma_f^2)$$

avec $\sigma_f^2 := \mu(|\sigma^* \nabla G_f|^2)$.

1.3.2. *Convergence des martingales.* Un cas particulier de (62) donne

$$(67) \quad \frac{\langle M^g \rangle_T}{T} \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} \mu(|g|^2) < +\infty.$$

Les hypothèses du théorème de la limite centrale pour les martingales sont donc vérifiées (voir Kutoyants[54] ou Basawa et Rao [8]) et

$$(68) \quad \frac{M_T^g}{\sqrt{T}} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \mu(|g|^2)).$$

1.3.3. *Convergence des martingales normalisées.* Une conséquence immédiate de (67) est

$$\langle M^g \rangle_T \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} +\infty$$

qui est une condition suffisante (voir Liptser et Shirayev [64]) pour établir la loi des grands nombres pour les martingales normalisées

$$(69) \quad \frac{M_T^g}{\langle M^g \rangle_T} \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0.$$

En combinant (67) et (68) nous obtenons une vitesse de convergence associée

$$(70) \quad \sqrt{T} \frac{M_T^g}{\langle M^g \rangle_T} \xrightarrow[T \rightarrow +\infty]{\text{Loi}} \mathcal{N}\left(0, \frac{1}{\mu(|g|^2)}\right).$$

1.4. Une remarque concernant l'hypothèse (H4). L'hypothèse (H4) est l'équivalent, dans notre contexte, de la condition de Cramer pour les développements d'Edgeworth dans le cas des variables aléatoires indépendantes et identiquement distribuées (voir [41]). Bien que nous ne soyons pas jusqu'ici parvenu à le prouver, nous pensons que cette hypothèse pourrait être déduite des seules hypothèses (H1)–(H3). A notre connaissance, ce résultat n'existe dans la littérature que pour le processus d'Ornstein-Uhlenbeck et la fonction identité ([46] lemmes 3 et 3'). Les figures 1 à 4 présentent quelques résultats numériques illustrant notre conjecture.

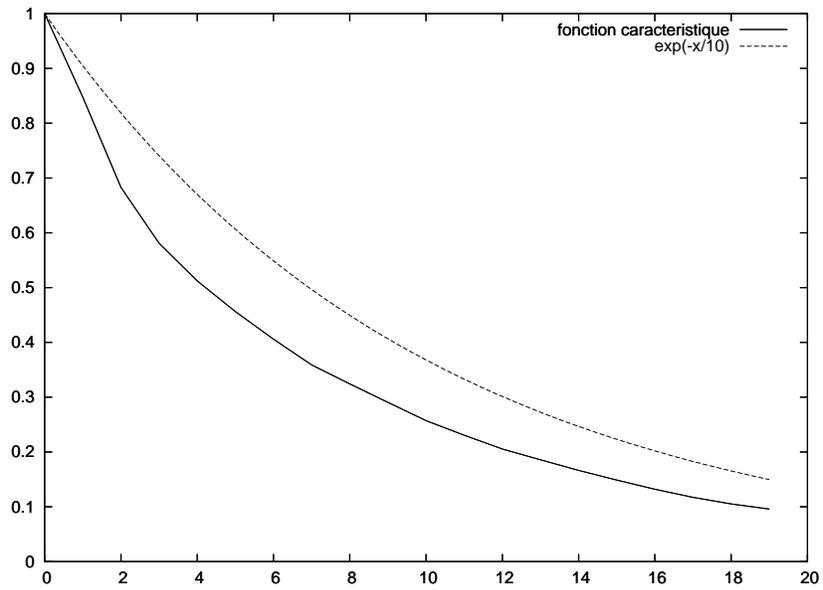


FIG. 1. Convergence de $\left| \mathbb{E} \left[\exp \left(i \int_0^T X_s dW_s \right) \right] \right|$ avec X solution de

$$X_t = -0.1 \int_0^t X_s ds + W_t$$

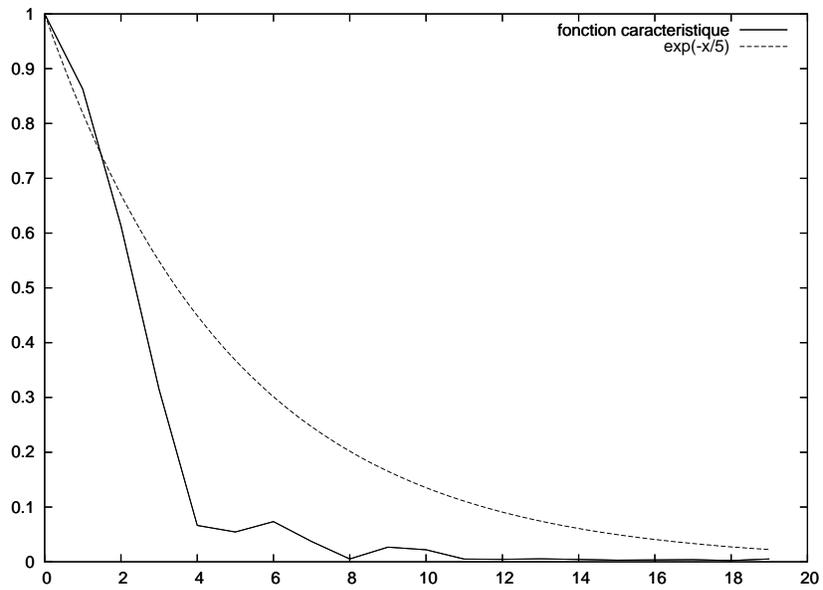


FIG. 2. Convergence de $\left| \mathbb{E} \left[\exp \left(i \int_0^T X_s^2 dW_s \right) \right] \right|$ avec X solution de

$$X_t = -0.1 \int_0^t X_s ds + W_t$$

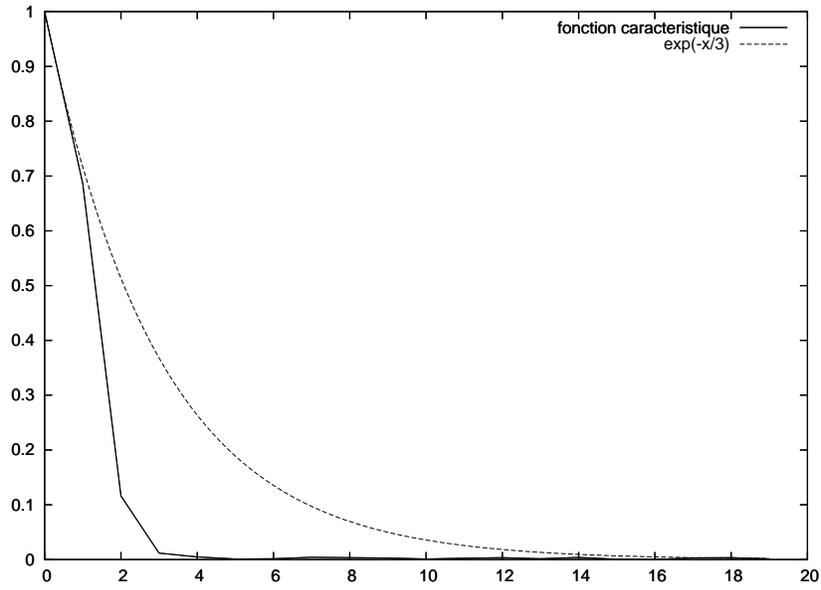


FIG. 3. Convergence de $\left| \mathbb{E} \left[\exp \left(i \int_0^T X_s^2 dW_s \right) \right] \right|$ avec X solution de

$$X_t = \int_0^t \tanh(X_s) ds + W_t$$

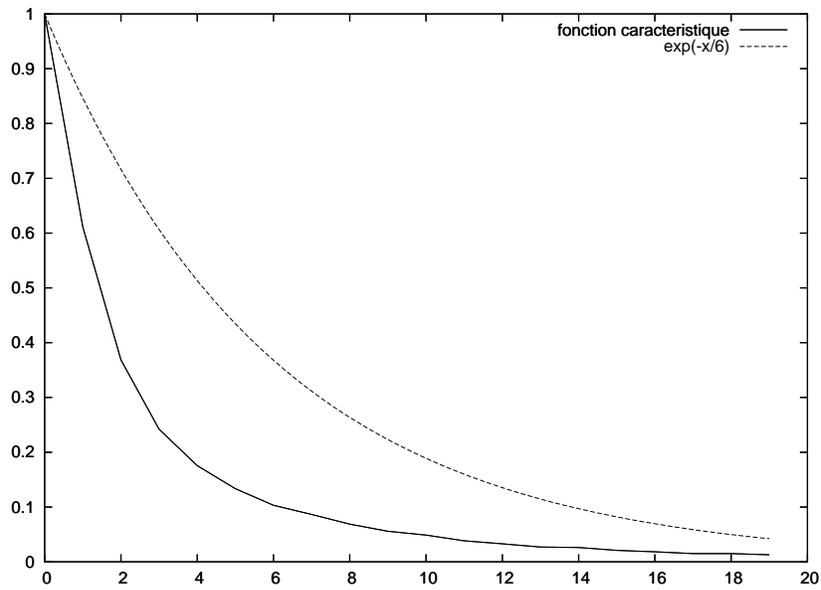


FIG. 4. Convergence de $\left| \mathbb{E} \left[\exp \left(i \int_0^T \frac{dW_s}{X_s} \right) \right] \right|$ avec X solution de

$$X_t = 1.0 + 0.1 \int_0^t (1.0 - X_s) ds + \int_0^t \sqrt{X_s} dW_s$$

1.5. Organisation du chapitre. L'objectif de ce chapitre est de préciser les résultats de convergence des équations (66), (68) et (70) en établissant un développement d'Edgeworth à l'ordre 1 pour les fonctions de répartition des différentes variables considérées.

Ce résultat sera établi dans un premier temps pour les martingales M^g en adaptant la preuve traditionnelle des développements d'Edgeworth dans le cas de variables indépendantes et identiquement distribuées due à Feller [28]. Cette preuve repose sur le contrôle de la différence entre la fonction caractéristique de la variable considérée et la fonction caractéristique de la loi gaussienne standard. La preuve se décompose en deux étapes. Loin de l'origine, le contrôle repose sur un développement de Taylor de la fonction caractéristique de la martingale M^g . Ce développement fait naturellement apparaître les moments de la martingale. Pour établir un développement d'Edgeworth à l'ordre 1, nous devons contrôler les trois premiers moments de M^g . C'est l'objet de la première partie de la section 2. Le contrôle de la différence entre les deux fonctions caractéristiques autour de l'origine repose sur la convergence exponentielle de ces deux fonctions vers 0. Ceci sera démontré au cours de la preuve du développement d'Edgeworth, dans la seconde partie de la section 2.

La preuve de la section 2 nous servira de patron pour étendre le résultat obtenu pour les martingales au cas des martingales normalisées et des moyennes empiriques. Pour ces deux preuves, nous nous contenterons donc de vérifier les éléments clés de la preuve de la section 2 ; la convergence exponentielle des fonctions caractéristiques vers 0 et les développements asymptotiques des trois premiers moments.

La section 3 présente un développement d'Edgeworth à l'ordre 1 pour les martingales normalisées et une application en statistique des processus.

La section 4 présente un développement d'Edgeworth à l'ordre 1 pour les moyennes empiriques. Pour établir ce résultat, nous avons alors besoin d'un contrôle sur la vitesse de convergence du gradient du semi-groupe p_f vers 0. Nous démontrons – sous des hypothèses renforcées – qu'il est possible d'obtenir une vitesse de convergence polynomiale. Une application du développement d'Edgeworth en simulation ergodique est également présentée.

Enfin, il est connu qu'un corollaire des développements d'Edgeworth sont les développements asymptotiques de quantiles ou développements de Cornish-Fisher. Nous rappelons ce résultat et présentons une application au bootstrap des estimateurs de maximum de vraisemblance de processus de diffusion dans la section 5 de ce chapitre.

2. Développement d'Edgeworth pour les martingales locales continues

Pour la fin de ce chapitre, Φ désignera la fonction de répartition de la loi normale centrée réduite et φ , sa densité.

2.1. Résultat principal. Le résultat suivant établit un développement asymptotique de la fonction de répartition d'une intégrale stochastique.

THÉORÈME 2.1.1. *Soit h_2 , le second polynôme de Hermite ($h_2(y) = y^2 - 1$). Sous les hypothèses (H1)–(H4),*

$$\mathbb{P}_x \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \leq y \right) = \Phi(y) - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2}) h_2(y)}{2\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} \varphi(y) + o\left(\frac{1}{\sqrt{T}}\right)$$

uniformément en y .

Nous retrouvons ainsi un résultats de Yoshida [87]. L'approche de Yoshida [87] repose sur l'étude de la régularité au sens de Malliavin de fonctionnelles définies sur l'espace de Wiener. Cette approche permet à l'auteur d'établir des résultats très généraux pour des martingales qui ne sont pas nécessairement des intégrales stochastiques du type M^g . Cette grande généralité se paie par des hypothèses parfois difficiles à vérifier. Notre cadre est plus spécifique puisqu'il ne concerne que les martingales locales de la famille M^g . En contrepartie, nous n'utilisons que des hypothèses très simples à vérifier et le cadre ergodique permet une démonstration classique de ce résultat fondée sur le développement asymptotiques des moments de martingales. En ce sens, notre approche est à relier à celle de Mykland [70].

2.2. Développement des moments. Pour établir notre développement d'Edgeworth, nous adapterons la preuve classique de Feller [28]. Pour ce faire, nous avons besoin de contrôler précisément les trois premiers moments des martingales M^g . C'est l'objet de cette section.

2.2.1. *Développement de l'espérance.* L'espérance ne pose évidemment aucun problème

$$(71) \quad \mathbb{E} \left[\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right] = 0.$$

2.2.2. *Développement de la variance.* Nous avons

$$\begin{aligned} \mathbb{V}\text{ar} \left[\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right] &= \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^2 \right] \text{ d'après (71)} \\ &= 1 + \frac{1}{\mu(|g|^2)T} \mathbb{E} [\langle M^g \rangle_T - \mu(|g|^2)T] \text{ par isométrie} \\ &= 1 + \frac{1}{\mu(|g|^2)T} \mathbb{E} \left[M_T^{\sigma^* \nabla G_{|g|^2}} - G_{|g|^2}(X_T) + G_{|g|^2}(x) \right] \text{ en utilisant (65)}. \end{aligned}$$

Or, la fonction $G_{|g|^2}$ est centrée, satisfait (H3) et donc, d'après la proposition 1.2.2, pour tout $k > 0$,

$$\exists s \in \mathbb{N}, \exists \Gamma > 0, |\mathbb{E} [G_{|g|^2}(X_T)]| \leq \Gamma(1 + |x|^s)(1 + T)^{-(k+1)}.$$

Par conséquent,

$$(72) \quad \mathbb{V}\text{ar} \left[\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right] = 1 + \frac{G_{|g|^2}(x)}{\mu(|g|^2)T} + O\left(\frac{1}{T^k}\right), \forall k \geq 1.$$

2.2.3. *Développement du troisième moment.* Posons

$$\begin{aligned}
\kappa_3(T) &:= \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \mathbb{E} \left[\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right] \right)^3 \right] \\
&= \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^3 \right] \text{ d'après (71).}
\end{aligned}$$

Rappelons que la puissance troisième de Wick

$$M_T^{g\circ 3} := (M_T^g)^3 - 3M_T^g \langle M^g \rangle_T$$

est une martingale (ce qui se vérifie aisément par une intégration par parties). Notre moment d'ordre 3 s'écrit donc

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^3 \right] &= \frac{3}{(\mu(|g|^2)T)^{\frac{3}{2}}} \mathbb{E} [M_T^g \langle M^g \rangle_T] \\
&= \frac{3}{(\mu(|g|^2)T)^{\frac{3}{2}}} \mathbb{E} [M_T^g (\langle M^g \rangle_T - \mu(|g|^2)T)] \\
&= \frac{3}{(\mu(|g|^2)T)^{\frac{3}{2}}} \mathbb{E} \left[M_T^g \left(M_T^{\sigma^* \nabla G_{|g|^2}} - G_{|g|^2}(X_T) + G_{|g|^2}(x) \right) \right]
\end{aligned}$$

en utilisant (65) avec la fonction $|g|^2$ qui vérifie (H3). En appliquant maintenant la propriété d'isométrie,

$$\begin{aligned}
\kappa_3(T) &= \frac{3}{(\mu(|g|^2)T)^{\frac{3}{2}}} \left(\mathbb{E} \left[\int_0^T (g \cdot \sigma^* \nabla G_{|g|^2})(X_s) ds \right] - \mathbb{E} [M_T^g G_{|g|^2}(X_T)] \right) \\
&= \frac{3\mu (g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} + \frac{3 \mathbb{E} [G_{g \cdot \sigma^* \nabla G_{|g|^2}}(x) - G_{g \cdot \sigma^* \nabla G_{|g|^2}}(X_T)]}{(\mu(|g|^2)T)^{\frac{3}{2}}} - \frac{3 \mathbb{E} [M_T^g G_{|g|^2}(X_T)]}{(\mu(|g|^2)T)^{\frac{3}{2}}}
\end{aligned}$$

en faisant une nouvelle fois appel à la représentation (65) avec la fonction $g \cdot \sigma^* \nabla G_{|g|^2}$ dont il est facile de vérifier qu'elle satisfait (H3). Le dernier terme se contrôle aisément car

$$\begin{aligned}
\frac{1}{T^{\frac{3}{2}}} |\mathbb{E} [M_T^g G_{|g|^2}(X_T)]| &\leq \frac{1}{T^{\frac{3}{2}}} \sqrt{\mathbb{E} [\langle M^g \rangle_T]} \sqrt{\mathbb{E} [G_{|g|^2}^2(X_T)]} \\
&= O\left(\frac{1}{T}\right).
\end{aligned}$$

Finalement, nous obtenons le développement

$$(73) \quad \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^3 \right] = \frac{3\mu (g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} + O\left(\frac{1}{T}\right).$$

2.3. Preuve du théorème 2.1.1. Rappelons le célèbre lemme d'Esseen (voir Petrov [73]) qui sera à la base de notre démonstration.

LEMME 2.3.1. Soit F , une fonction croissante, H une fonction à valeurs réelles, dérivable et à variation bornée, et f et h , leurs transformées de Fourier-Stieltjes respectives. Supposons que $F(-\infty) = H(-\infty)$ et $F(+\infty) = H(+\infty)$. Soit Ω , un nombre positif arbitraire. Supposons également que $\|H'\|_\infty \leq C$. Alors, pour tout $\beta > \frac{1}{2\pi}$, nous avons

$$\sup_{y \in \mathbb{R}} |F(y) - H(y)| \leq \beta \int_{-\Omega}^{+\Omega} \left| \frac{f(\omega) - h(\omega)}{\omega} \right| d\omega + r(\beta) \frac{C}{\Omega}$$

où $r(\beta)$ est une constante positive qui ne dépend que de β .

Passons maintenant à la preuve du théorème.

PREUVE DU THÉORÈME 2.1.1. Nous posons

$$F_T(y) := \mathbb{P}_x \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \leq y \right)$$

et

$$\Psi_T(y) := \Phi(y) - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2}) h_2(y)}{2\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} \varphi(y).$$

Les transformées de Fourier-Stieltjes respectives de ces deux fonctions sont définies par

$$\hat{F}_T(\omega) := \mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right) \right]$$

et

$$\begin{aligned} \hat{\Psi}_T(\omega) &:= \int_{\mathbb{R}} e^{i\omega y} d\Psi_T(y) \\ &= e^{-\frac{\omega^2}{2}} \left(1 + \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{2\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} (i\omega)^3 \right) \end{aligned}$$

Soient maintenant $\varepsilon > 0$ et $A > \frac{24}{\varepsilon\pi} \|\Psi_T'\|_\infty$.

Première étape : contrôle de la fonction caractéristique loin de l'origine

Pour $\Omega = A\sqrt{T}$, le lemme 2.3.1 nous donne la majoration

$$(74) \quad \sup_{y \in \mathbb{R}} |F_T(y) - \Psi_T(y)| \leq \frac{1}{\pi} \int_{-A\sqrt{T}}^{A\sqrt{T}} \left| \frac{\hat{F}_T(\omega) - \hat{\Psi}_T(\omega)}{\omega} \right| d\omega + \frac{\varepsilon}{\sqrt{T}}.$$

Nous allons montrer que l'intégrale figurant dans cette inégalité est d'ordre $o\left(\frac{1}{\sqrt{T}}\right)$. Choisissons un réel $0 < B < A$ et partageons le domaine d'intégration en $[0, B\sqrt{T}]$ et $[B\sqrt{T}, A\sqrt{T}]$. Sur le domaine $B\sqrt{T} \leq |\omega| \leq A\sqrt{T}$, l'intégrale en question est majorée par

$$\frac{1}{\pi B} \int_{B\sqrt{T} \leq |\omega| \leq A\sqrt{T}} \left| \hat{F}_T(\omega\sqrt{T}) \right| d\omega + \frac{1}{\pi B} \int_{B\sqrt{T} \leq |\omega| \leq A\sqrt{T}} \left| \hat{\Psi}_T(\omega\sqrt{T}) \right| d\omega.$$

Il est connu (voir toujours Feller [28]) que le second membre converge exponentiellement vite vers 0. Nous considérons maintenant l'intégrande du premier membre. Celui-ci s'écrit également :

$$\left| \hat{F}_T(\omega\sqrt{T}) \right| = \left| \mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)}} \right) \right] \right|.$$

Or, l'hypothèse (H4) nous assure de l'existence de réels $T_0 > 0$, $\omega_0 > 0$ et $\gamma > 0$ tels que, pour tout $|\omega| > \omega_0$,

$$|\mathbb{E} [\exp (i\omega M_T^g)]| \leq e^{-\gamma T}, \forall T \geq T_0.$$

Ceci conclut la première partie de cette démonstration.

Seconde étape : contrôle de la fonction caractéristique autour de l'origine

Supposons maintenant $|\omega| < B\sqrt{T}$. L'intégrale que nous cherchons à majorer s'écrit alors

$$(75) \quad \int_{-B\sqrt{T}}^{B\sqrt{T}} e^{-\frac{\omega^2}{2}} \left| \frac{e^{\frac{\omega^2}{2}} \mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right) \right] - 1 - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{2\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} (i\omega)^3}{\omega} \right| d\omega.$$

Nous allons utiliser l'inégalité

$$|e^\alpha - 1 - \beta| \leq (|\alpha - \beta| + \frac{\beta^2}{2})e^\gamma$$

avec $\gamma \geq (|\alpha| \vee |\beta|)$ (voir Feller [28]). Nous choisissons évidemment

$$\alpha = \frac{\omega^2}{2} + \ln \mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right) \right]$$

et

$$\beta = \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{2\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} (i\omega)^3.$$

Or, comme M_T^g admet un moment d'ordre 3 fini, il existe un petit voisinage autour de 0 dans lequel la fonction caractéristique $\mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right) \right]$ est trois fois dérivable. Etant donné $\varepsilon > 0$ fixé, il est donc possible de choisir le réel B de telle sorte que, d'une part,

$$|\beta| \leq \frac{\omega^2}{4}$$

et que, d'autre part, la différence entre la fonction caractéristique et ses développements de Taylor aux ordres 2 et 3 soit majorée par

$$\left| \ln \mathbb{E} \left[\exp \left(i\omega \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right) \right] + \frac{\omega^2}{2} \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^2 \right] \right| \leq \frac{\omega^2}{4} \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^2 \right]$$

et

$$\left| \ln \mathbb{E} \left[\exp \left(i\omega \frac{M_T}{\sqrt{\mu(|g|^2)T}} \right) \right] + \frac{\omega^2}{2} \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^2 \right] - \frac{(i\omega)^3}{6} \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^3 \right] \right| \leq \varepsilon |\omega|^3 \left| \mathbb{E} \left[\left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \right)^3 \right] \right|.$$

En utilisant les estimations (72) et (73) et le fait que $|\omega| \leq B\sqrt{T}$ pour B arbitrairement petit, nous déduisons de ces deux expressions que

$$|\alpha| \leq \frac{\omega^2}{4} + \varepsilon$$

et

$$|\alpha - \beta| \leq \frac{\varepsilon |\omega|^3}{\sqrt{T}} + (\omega^2 + |\omega|^3) O\left(\frac{1}{T}\right).$$

Par ailleurs, du fait des majorations de α et β , nous posons évidemment

$$\gamma = \frac{\omega^2}{4} + \varepsilon.$$

Nous vérifions alors que l'intégrale (75) est majorée par

$$\begin{aligned} \int_{-B\sqrt{T}}^{B\sqrt{T}} e^{-\frac{\omega^2}{4} + \varepsilon} \left(\frac{\varepsilon |\omega|^3}{\sqrt{T}} + (\omega^2 + |\omega|^3) O\left(\frac{1}{T}\right) + \frac{\mu (g \cdot \sigma^* \nabla G_{|g|^2})^2}{8\mu(|g|^2)^3 T} \omega^5 \right) \frac{d\omega}{|\omega|} \\ \leq \frac{\varepsilon e^\varepsilon}{\sqrt{T}} \int_{-\infty}^{+\infty} \omega^2 e^{-\frac{\omega^2}{4}} d\omega + O\left(\frac{1}{T}\right). \end{aligned}$$

Comme ε est arbitraire, nous obtenons bien le résultat annoncé. \square

Nous allons maintenant étendre ce résultat à d'autres fonctionnelles. Pour ce faire, nous allons exprimer celles-ci sous la forme de développements asymptotiques dont le terme dominant est une martingale. La preuve précédente fait apparaître les deux éléments permettant d'obtenir les développements d'Edgeworth. La première partie de la preuve repose sur la vitesse de convergence de la fonction de caractéristique de notre fonctionnelle. Nous avons vu que cette vitesse est exponentielle dans le cas des martingales. Par la suite cette vitesse sera conservée du fait de la présence d'une martingale en tant que terme principal de nos développements asymptotiques. La deuxième partie de la preuve est plus algébrique. Il s'agit en effet d'identifier les coefficients du développement grâce aux développements asymptotiques des moments.

3. Développement d'Edgeworth pour les martingales locales continues normalisées

3.1. Résultat principal. Nous étendons maintenant le théorème 2.1.1 au cas des martingales normalisées par leur processus croissant.

THÉORÈME 3.1.1. *Sous les hypothèses (H1)–(H4), si la valeur initiale $X_0 = x$ appartient au domaine de g et si $\langle M^g \rangle_T$ admet des moments inverses de tous ordres finis pour T suffisamment grand, alors*

$$\begin{aligned} \mathbb{P}_x \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) &= \Phi(y) - \frac{\varphi(y)}{\sqrt{\mu(|g|^2)T}} \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2}) h_2(y)}{2\mu(|g|^2)} \\ &\quad + \frac{\varphi(y)}{\sqrt{\mu(|g|^2)T}} \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)} + o\left(\frac{1}{\sqrt{T}}\right) \end{aligned}$$

uniformément en y .

Notons que le biais de notre variable apparaît au travers du terme $\frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)}$. Supposer X_0 dans le domaine de g permet de s'assurer que $\langle M^g \rangle_T > 0$ pour tout $T > 0$.

DÉMONSTRATION. Nous allons suivre la même démarche que lors de la preuve du théorème 2.1.1.

Préliminaires

Nous allons appliquer la méthode du delta (voir Hall [41]). Un développement de Taylor avec reste intégral à l'ordre 1 nous donne

$$\begin{aligned} \sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} &= \frac{\frac{M_T^g}{\sqrt{\mu(|g|^2)T}}}{\left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1\right) + 1} \\ &= \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1\right) \\ &\quad + 2 \underbrace{\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1\right)^2 \int_0^1 \frac{(1-t) dt}{\left(1 + t \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1\right)\right)^3}}_{=:\Delta_2(T)}. \end{aligned}$$

Nous commençons par développer le second terme.

$$\begin{aligned}
\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1 \right) &= \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{M_T^{\sigma^* \nabla G_{|g|^2}}}{\mu(|g|^2)T} + \frac{G_{|g|^2}(x)}{\mu(|g|^2)T} - \frac{G_{|g|^2}(X_T)}{\mu(|g|^2)T} \right) \\
&= \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{G_{|g|^2}(x)}{\mu(|g|^2)T} - \frac{G_{|g|^2}(X_T)}{\mu(|g|^2)T} \right) \\
&\quad + \frac{1}{(\mu(|g|^2)T)^{\frac{3}{2}}} \left(M_T^g M_T^{\sigma^* \nabla G_{|g|^2}} - \langle M^g, M^{\sigma^* \nabla G_{|g|^2}} \rangle_T \right) \\
&\quad + \frac{1}{(\mu(|g|^2)T)^{\frac{3}{2}}} \langle M^g, M^{\sigma^* \nabla G_{|g|^2}} \rangle_T \\
&= \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{G_{|g|^2}(x)}{\mu(|g|^2)T} - \frac{G_{|g|^2}(X_T)}{\mu(|g|^2)T} \right) \\
&\quad + \frac{1}{(\mu(|g|^2)T)^{\frac{3}{2}}} \left(M_T^g M_T^{\sigma^* \nabla G_{|g|^2}} - \langle M^g, M^{\sigma^* \nabla G_{|g|^2}} \rangle_T \right) \\
&\quad + \frac{1}{\sqrt{T} \mu(|g|^2)^{\frac{3}{2}}} \left(\frac{1}{T} \int_0^T (g \cdot \sigma^* \nabla G_{|g|^2})(X_s) ds - \mu(g \cdot \sigma^* \nabla G_{|g|^2}) \right) \\
&\quad + \frac{1}{\sqrt{T} \mu(|g|^2)^{\frac{3}{2}}} \mu(g \cdot \sigma^* \nabla G_{|g|^2}).
\end{aligned}$$

Nous posons maintenant

$$\begin{aligned}
\Delta_1(T) &:= \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{G_{|g|^2}(x)}{\mu(|g|^2)T} - \frac{G_{|g|^2}(X_T)}{\mu(|g|^2)T} \right) \\
&\quad + \frac{1}{(\mu(|g|^2)T)^{\frac{3}{2}}} \left(M_T^g M_T^{\sigma^* \nabla G_{|g|^2}} - \langle M^g, M^{\sigma^* \nabla G_{|g|^2}} \rangle_T \right) \\
&\quad + \frac{1}{\sqrt{T} \mu(|g|^2)^{\frac{3}{2}}} \left(\frac{1}{T} \int_0^T (g \cdot \sigma^* \nabla G_{|g|^2})(X_s) ds - \mu(g \cdot \sigma^* \nabla G_{|g|^2}) \right),
\end{aligned}$$

de telle sorte que

$$(76) \quad \sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} = \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) + \Delta_2(T).$$

En utilisant le lemme 3.1.1 (voir également la preuve du théorème 1 de Sakamoto et Yoshida [79]) nous vérifions maintenant qu'il existe $C > 0$ et $\varepsilon > 0$ tels que

$$\mathbb{P} \left(|\Delta_2(T)| > CT^{-\frac{1}{2}-\varepsilon} \right) = o \left(\frac{1}{\sqrt{T}} \right).$$

Alors, d'une part,

$$\begin{aligned}
\mathbb{P} \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) &= \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) + \Delta_2(T) \leq y \right) \\
&\leq \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y + CT^{-\frac{1}{2}-\varepsilon} \right) \\
&\quad + \mathbb{P} \left(|\Delta_2(T)| > CT^{-\frac{1}{2}-\varepsilon} \right) \\
&= \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y \right) + o \left(\frac{1}{\sqrt{T}} \right),
\end{aligned}$$

et, d'autre part, comme

$$\begin{aligned}
&\mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y - CT^{-\frac{1}{2}-\varepsilon} \right) \\
\leq \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y - \Delta_2(T), |\Delta_2(T)| \leq CT^{-\frac{1}{2}-\varepsilon} \right) \\
&\quad + \mathbb{P} \left(|\Delta_2(T)| > CT^{-\frac{1}{2}-\varepsilon} \right) \\
\leq \mathbb{P} \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) + \mathbb{P} \left(|\Delta_2(T)| > CT^{-\frac{1}{2}-\varepsilon} \right),
\end{aligned}$$

nous avons également,

$$\begin{aligned}
\mathbb{P} \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) &\geq \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y - CT^{-\frac{1}{2}-\varepsilon} \right) \\
&\quad - \mathbb{P} \left(|\Delta_2(T)| > CT^{-\frac{1}{2}-\varepsilon} \right) \\
&\geq \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y \right) \\
&\quad + o \left(\frac{1}{\sqrt{T}} \right).
\end{aligned}$$

Le fait que

$$\begin{aligned}
&\mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y \pm CT^{-\frac{1}{2}-\varepsilon} \right) \\
= \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y \right) + o \left(\frac{1}{\sqrt{T}} \right)
\end{aligned}$$

s'obtient à partir d'un développement de Taylor dans le développement d'Edgeworth du théorème 2.1.1. Nous avons donc montré que

$$\mathbb{P} \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) = \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} - \Delta_1(T) \leq y \right) + o \left(\frac{1}{\sqrt{T}} \right).$$

De même, il est immédiat de vérifier à partir de l'expression de $\Delta_1(T)$ que

$$\mathbb{E} [|\Delta_1(T)|^p] = O \left(\frac{1}{T^p} \right).$$

Ainsi, en suivant le même raisonnement que précédemment, nous établissons que

$$\mathbb{P} \left(\sqrt{\mu(|g|^2)T} \frac{M_T^g}{\langle M^g \rangle_T} \leq y \right) = \mathbb{P} \left(\frac{M_T^g}{\sqrt{\mu(|g|^2)T}} - \frac{\mu(g \cdot \sigma^* \nabla G_{|g|^2})}{\mu(|g|^2)^{\frac{3}{2}} \sqrt{T}} \leq y \right) + o \left(\frac{1}{\sqrt{T}} \right).$$

Il ne reste plus qu'à procéder à un développement de Taylor dans l'expression du théorème 2.1.1 pour établir le résultat annoncé. □

Nous présentons maintenant les deux lemmes mentionnés au cours de la preuve précédente.

LEMME 3.1.1. Avec les hypothèses du théorème 3.1.1 et les notations de sa preuve, pour tout $p \geq 1$,

$$\mathbb{E} [|\Delta_2(T)|^p] = O \left(\frac{1}{T^p} \right).$$

DÉMONSTRATION. Rappelons que le terme $\Delta_2(T)$ s'écrit

$$\begin{aligned} \Delta_2(T) &= 2 \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1 \right)^2 \int_0^1 \frac{(1-t) dt}{\left(1 + t \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1 \right) \right)^3} \\ &= 2 \sqrt{\mu(|g|^2)T} M_T^g (\langle M^g \rangle_T - \mu(|g|^2)T)^2 \int_0^1 \frac{(1-t) dt}{(\mu(|g|^2)T + t(\langle M^g \rangle_T - \mu(|g|^2)T))^3}. \end{aligned}$$

Pour poursuivre la démonstration, nous avons à étudier la fonction

$$\varphi : t \mapsto \frac{(1-t)}{(\mu(|g|^2)T + t(\langle M^g \rangle_T - \mu(|g|^2)T))^3}$$

sur l'intervalle $[0, 1]$. Vérifions tout d'abord qu'elle est bien définie. Le dénominateur ne peut s'annuler qu'en la valeur $t_0 := \frac{\mu(|g|^2)T}{\mu(|g|^2)T - \langle M^g \rangle_T}$. Or, nous avons la suite d'équivalences

$$\begin{aligned}
& 0 \leq t_0 \leq 1 \\
& \Leftrightarrow 0 \leq \frac{\mu(|g|^2)}{\mu(|g|^2) - \frac{1}{T}\langle M^g \rangle_T} \leq 1 \\
& \Leftrightarrow 1 \leq \frac{\mu(|g|^2) - \frac{1}{T}\langle M^g \rangle_T}{\mu(|g|^2)} \\
& \Leftrightarrow \mu(|g|^2) \leq \mu(|g|^2) - \frac{1}{T}\langle M^g \rangle_T \\
& \Leftrightarrow \frac{1}{T}\langle M^g \rangle_T \leq 0.
\end{aligned}$$

Comme cet événement est de probabilité nulle par hypothèse (nous avons en effet supposé que X_0 est dans le domaine de f), la fonction φ est presque sûrement continue sur le compact $[0, 1]$. En particulier, elle est bornée sur ce compact. Par ailleurs, une étude des variations de φ permet de vérifier qu'elle ne comporte sur \mathbb{R} qu'un unique extremum local, en dehors de celui éventuellement atteint en t_0 . Cet extremum local est atteint en $t_1 := \frac{3}{2} - \frac{\mu(|g|^2)T}{2(\mu(|g|^2)T - \langle M^g \rangle_T)}$. Deux cas de figure sont alors envisageables :

– Soit $t_1 \in [0, 1]$, auquel cas la fonction φ est majorée sur $[0, 1]$ par

$$\frac{2}{27} \frac{1}{\langle M^g \rangle_T^2 (\mu(|g|^2)T - \langle M^g \rangle_T)}.$$

Ceci survient sur l'événement $\left\{ \frac{\langle M^g \rangle_T}{T} \leq \frac{2}{3}\mu(|g|^2) \right\}$.

– Soit $t_1 \notin [0, 1]$, auquel cas l'extremum de φ sur $[0, 1]$ est atteint sur l'un des bords de l'intervalle. Il est facile de voir que ce bord est 0 et que l'extremum est alors $(\mu(|g|^2)T)^{-3}$.

Ceci survient sur l'événement complémentaire $\left\{ \frac{\langle M^g \rangle_T}{T} \geq \frac{2}{3}\mu(|g|^2) \right\}$.

Revenons maintenant au terme de reste, que nous décomposons sur les deux événements considérés précédemment.

$$\begin{aligned}
\mathbb{E}[|\Delta_2(T)|^p] &= 2^p \mathbb{E} \left[\left| \sqrt{\mu(|g|^2)T} M_T^g (\langle M^g \rangle_T - \mu(|g|^2)T)^2 \int_0^1 \frac{(1-t) dt}{(\mu(|g|^2)T + t(\langle M^g \rangle_T - \mu(|g|^2)T))^3} \right|^p \right] \\
&\leq \frac{2^p}{(\mu(|g|^2)T)^{3p}} \mathbb{E} \left[\left| \sqrt{\mu(|g|^2)T} M_T^g (\langle M^g \rangle_T - \mu(|g|^2)T)^2 \right|^p \mathbb{1}_{\left(\frac{\langle M^g \rangle_T}{T} \geq \frac{2}{3}\mu(|g|^2)\right)} \right] \\
&\quad + 6^p \mathbb{E} \left[\left| \frac{4\sqrt{\mu(|g|^2)T} (\mu(|g|^2)T - \langle M^g \rangle_T)}{27\langle M^g \rangle_T} \right|^p \mathbb{1}_{\left(\frac{\langle M^g \rangle_T}{T} \leq \frac{2}{3}\mu(|g|^2)\right)} \right] \\
&\leq 2^p \mathbb{E} \left[\left| \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1 \right)^2 \right|^p \right] \\
&\quad + 6^p \sqrt{\mathbb{E} \left[\left(\frac{4\sqrt{\mu(|g|^2)T} (\mu(|g|^2)T - \langle M^g \rangle_T)}{27\langle M^g \rangle_T} \right)^{2p} \right]} \sqrt{\mathbb{P} \left(\frac{\langle M^g \rangle_T}{T} \leq \frac{2}{3}\mu(|g|^2) \right)}.
\end{aligned}$$

En utilisant à nouveau la représentation de Poisson et l'inégalité de Cauchy Schwarz, nous établissons que le premier terme est d'ordre

$$\mathbb{E} \left[\left| \frac{M_T^g}{\sqrt{\mu(|g|^2)T}} \left(\frac{\langle M^g \rangle_T}{\mu(|g|^2)T} - 1 \right) \right|^{2p} \right] = O \left(\frac{1}{T^p} \right).$$

Nous utilisons maintenant le lemme 3.1.2 pour établir que le second terme est d'ordre $O \left(\frac{1}{T^p} \right)$. En effet,

$$\mathbb{P} \left(\frac{\langle M^g \rangle_T}{T} \leq \frac{2}{3} \mu(|g|^2) \right) = O \left(\frac{1}{T^n} \right), \forall n > 2p,$$

et du fait de nos hypothèses,

$$\mathbb{E} \left[\left(\frac{4\sqrt{\mu(|g|^2)T} (\mu(|g|^2)T - \langle M^g \rangle_T)}{27\langle M_g \rangle_T} \right)^{2p} \right] = O(T^p).$$

Ainsi,

$$\mathbb{E} [|\Delta_2(T)|^p] = O \left(\frac{1}{T^p} \right).$$

et notre lemme est démontré. \square

Le dernier lemme énonce une borne sur la probabilité de déviation intervenue au cours de la démonstration précédente. Nous savons que, sous des hypothèses plus fortes que les nôtres, il est possible d'obtenir des bornes exponentielles pour les probabilités de déviation (voir Lezaud[60] ou Kontoyiannis et Meyn [50]). Nous ne sommes pas parvenus à obtenir des telles bornes dans le cadre que nous avons choisi. Toutefois, pour notre propos, une borne polynomiale est suffisante.

LEMME 3.1.2. Pour tout $\eta > 0$:

$$\mathbb{P}_x (|\langle M^g \rangle_T - T\mu(|g|^2)| > \eta T) = O \left(\frac{1}{T^n} \right), \forall n > 0.$$

DÉMONSTRATION. Remarquons avant tout que, sous les hypothèses (H1)–(H4), un résultat de E. Pardoux et A. Yu. Veretennikov [72] nous assure que, si la fonction $|g|^2$ vérifie (H3) alors la fonction $G_{|g|^2}$ et sa dérivée sont à croissance polynomiale. L'utilisation de l'inégalité de Markov et de la représentation de Poisson (65) nous donnent alors :

$$\begin{aligned} & \mathbb{P}_x (|\langle M^g \rangle_T - T\mu(|g|^2)| > \eta T) \\ & \leq \frac{\mathbb{E}_x \left[(\langle M^g \rangle_T - T\mu(|g|^2))^{2n} \right]}{(\eta T)^{2n}} \\ & \leq \frac{\mathbb{E}_x \left[\left(\int_0^T \sigma^*(X_s) \nabla G_{|g|^2}(X_s) \cdot dW_s - (G_{|g|^2}(X_T) - G_{|g|^2}(x)) \right)^{2n} \right]}{(\eta T)^{2n}} \\ & \leq \frac{\mathbb{E}_x \left[\left(\left| \int_0^T \sigma^*(X_s) \nabla G_{|g|^2}(X_s) \cdot dW_s \right| + |G_{|g|^2}(X_T) - G_{|g|^2}(x)| \right)^{2n} \right]}{(\eta T)^{2n}}. \end{aligned}$$

Or, pour tous réels positifs a et b , et tout entier p , nous avons $(a + b)^p \leq 2^p (a^p + b^p)$. Par conséquent, notre inégalité devient :

$$\begin{aligned} \mathbb{P}_x \left(\left| \langle M^g \rangle_T - T\mu(|g|^2) \right| > \eta T \right) &\leq 2^{2n} \frac{\mathbb{E}_x \left[\left(\int_0^T \sigma^*(X_s) \nabla G_{|g|^2}(X_s) \cdot dW_s \right)^{2n} \right]}{(\eta T)^{2n}} \\ &+ 2^{2n} \frac{\mathbb{E}_x \left[(G_{|g|^2}(X_T) - G_{|g|^2}(x))^{2n} \right]}{(\eta T)^{2n}}. \end{aligned}$$

Nous utilisons maintenant l'inégalité de Burkholder-Davis-Gundy et le fait que les fonctions σ^* , $G_{|g|^2}$ et $\nabla G_{|g|^2}$ soient à croissance polynomiale et que les moments de X_t soient uniformément bornés pour tout $t \geq 0$. Il existe donc des constantes strictement positives c_n , β_n et \tilde{c}_n telles que

$$\mathbb{P}_x \left(\left| \langle M^g \rangle_T - T\mu(|g|^2) \right| > \eta T \right) \leq \frac{c_n T^n}{(\eta T)^{2n}} + \frac{\beta_n}{(\eta T)^{2n}}.$$

Par conséquent, nous obtenons

$$\mathbb{P}_x \left(\left| \langle M^g \rangle_T - T\mu(|g|^2) \right| > \eta T \right) \leq \frac{\tilde{c}_n}{T^n}.$$

Ceci nous donne le résultat annoncé. □

Nous allons maintenant appliquer ce résultat en statistique des processus. Notons toutefois que les martingales normalisées apparaissent dans d'autre contexte, par exemple lors de l'étude de la régularité de la densité des processus de diffusion (voir [20]).

3.2. Applications en statistique des processus.

3.2.1. *Développement d'Edgeworth.* Nous considérons maintenant un processus de diffusion unidimensionnel dont le coefficient de dérive dépend linéairement d'un paramètre θ , observé sur un intervalle de temps $[0, T]$:

$$dX_t = \theta b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = x, \quad 0 \leq t \leq T.$$

Il est possible de construire un estimateur de θ par la méthode du maximum de vraisemblance (voir Liptser et Shiryaev [64]). Soit donc la fonction de vraisemblance définie par

$$\lambda_T(\theta, X) := \exp \left(\int_0^T \frac{\theta b(X_t)}{\sigma^2(X_t)} dX_t - \frac{1}{2} \int_0^T \left(\frac{\theta b(X_t)}{\sigma(X_t)} \right)^2 dt \right)$$

L'estimateur de maximum de vraisemblance $\hat{\theta}_T$ est défini comme la valeur du paramètre θ qui maximise cette fonctionnelle. En règle générale, cet estimateur est construit en recherchant la valeur qui annule la dérivée logarithmique de la fonction de vraisemblance :

$$\frac{\partial}{\partial \theta} \ln \lambda_T(\theta, X) = \int_0^T \frac{b(X_t)}{\sigma^2(X_t)} dX_t - \theta \int_0^T \frac{b^2(X_t)}{\sigma^2(X_t)} dt.$$

Il est clair que

$$\hat{\theta}_T = \frac{\int_0^T \frac{b(X_t)}{\sigma^2(X_t)} dX_t}{\int_0^T \frac{b^2(X_t)}{\sigma^2(X_t)} dt}$$

et l'équation vérifiée par le processus X permet d'obtenir l'expression

$$\hat{\theta}_T - \theta = \frac{\int_0^T \frac{b(X_t)}{\sigma(X_t)} dW_t}{\int_0^T \left(\frac{b(X_t)}{\sigma(X_t)} \right)^2 dt}.$$

Il est donc clair d'après (69) et (70) que cette estimateur est fortement consistant et asymptotiquement normal, et une application directe du théorème 3.1.1 nous donne alors la

PROPOSITION 3.2.1. *Sous les hypothèses du théorème 3.1.1, nous avons le développement*

$$\begin{aligned} \mathbb{P}_x \left(\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right) T} (\hat{\theta}_T - \theta) \leq y \right) &= \Phi(y) - \frac{\varphi(y)}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right) T}} \left(\frac{\mu \left(bG'_{\frac{b^2}{\sigma^2}} \right) h_2(y)}{2\mu \left(\frac{b^2}{\sigma^2} \right)} \right) \\ &+ \frac{\varphi(y)}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right) T}} \frac{\mu \left(bG'_{\frac{b^2}{\sigma^2}} \right)}{\mu \left(\frac{b^2}{\sigma^2} \right)} + o \left(\frac{1}{\sqrt{T}} \right) \end{aligned}$$

uniformément en y .

Cette proposition généralise -à l'ordre 1- les résultats obtenus pour le processus d'Ornstein-Uhlenbeck par Bishwal et Bose [14] et Kakizawa [46]. Par ailleurs, nous retrouvons dans le cas particulier d'un modèle linéaire le résultat obtenu par Yoshida [87]. Des résultats d'ordre supérieur sont connus dans le cas particulier du processus d'Ornstein-Uhlenbeck. Citons entre autre Kakizawa [46], Bercu et Rouault [11], Florens-Landais et Pham [29] et Bishwal

et Bose [14]. Dans ce cas particulier, la plupart de ces résultats reposent sur l'expression explicite de la fonction caractéristique de $\int_0^t X_s^2 ds$ (voir Liptser et Shirayev [64]). Il est possible d'étendre sans difficulté ce résultat au cas où le coefficient de dérive ne dépend pas linéairement du paramètre θ (voir Yoshida [87]).

3.2.2. *Correction du biais.* Il n'est pas en générale possible de déduire des développements de moments du développement d'Edgeworth. Toutefois, une conséquence de la preuve du théorème 3.1.1 est le développement de l'espérance des martingales normalisées ; il suffit de prendre l'espérance dans l'équation (76) et d'utiliser les estimations sur $\Delta_1(T)$ et $\Delta_2(T)$. Appliqué au contexte de l'estimateur de maximum de vraisemblance, ce résultat implique l'existence d'une constante c indépendante du temps et telle que le biais de l'EMV se développe en

$$\mathbb{E} [\hat{\theta}_T - \theta] = \frac{c}{T} + O\left(\frac{1}{T^{\frac{3}{2}}}\right).$$

Si nous cherchons à corriger le biais de notre estimateur $\hat{\theta}_T$, il est alors possible d'appliquer une méthode de type interpolation de Romberg. En effet, définissons le nouvel estimateur

$$\tilde{\theta}_T := 2\hat{\theta}_T - \hat{\theta}_{\frac{T}{2}}$$

fondé sur l'observation du processus X sur les intervalles $[0, T]$ et $[0, \frac{T}{2}]$. Le biais de l'estimateur $\tilde{\theta}_T$ est d'ordre $O\left(\frac{1}{T^{\frac{3}{2}}}\right)$. En effet, ce biais s'écrit :

$$\begin{aligned} \mathbb{E} [\tilde{\theta}_T - \theta] &= 2\mathbb{E} [\hat{\theta}_T - \theta] - \mathbb{E} [\hat{\theta}_{\frac{T}{2}} - \theta] \\ &= 2\frac{c}{T} - \frac{c}{T/2} + O\left(\frac{1}{T^{\frac{3}{2}}}\right) \\ &= O\left(\frac{1}{T^{\frac{3}{2}}}\right). \end{aligned}$$

Nous avons donc gagné un ordre de convergence. La figure 5 illustre cette procédure sur l'exemple du processus d'Ornstein-Uhlenbeck. Mentionnons l'article d'E. Clément [23] où une méthode de correction du biais proche du bootstrap est présentée.

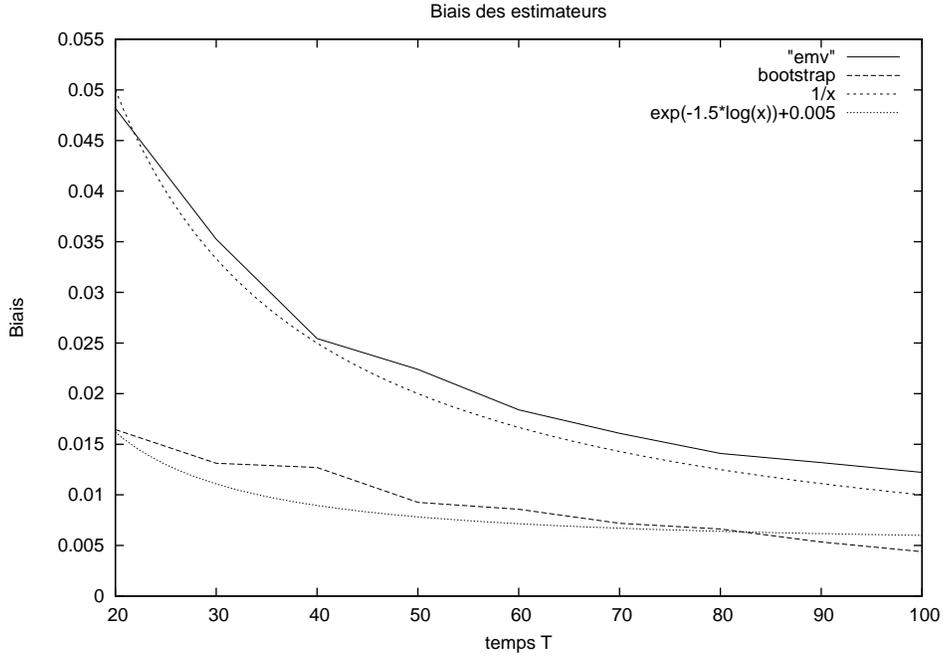


FIG. 5. Correction du biais pour l'EMV du processus d'Orstein-Uhlenbeck de paramètre $\theta = \frac{1}{2}$

4. Développement d'Edgeworth pour les moyennes empiriques

Pour cette dernière section, nous renforçons nos hypothèses initiales (H1)–(H4) et supposons de plus que

(H5) Les fonctions b et σ sont de classes C^∞ à dérivées de tous ordres bornées. La fonction σ est bornée.

(H6) La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est de classe C^∞ .

(H7) Il existe des réels $T_0 > 0$, $\omega_0 > 0$ et $\gamma > 0$ tels que, pour tout $|\omega| > \omega_0$,

$$\left| \mathbb{E} \left[\exp \left(i\omega \int_0^T \tilde{f}(X_s) ds \right) \right] \right| \leq e^{-\gamma T}, \forall T \geq T_0.$$

4.1. Résultat principal. Le résultat suivant est une extension du théorème 2.1.1 au cas des moyennes empiriques.

THÉORÈME 4.1.1. *Sous les hypothèses (H1)–(H7),*

$$\mathbb{P}_x \left(\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T \tilde{f}(X_s) ds \leq y \right) = \Phi(y) - \frac{\mu(\sigma^* \nabla G_f \cdot \sigma^* \nabla G_{|\sigma^* \nabla G_f|^2}) h_2(y)}{2\sigma_f^3 \sqrt{T}} \varphi(y) + \frac{G_f(x)}{\sqrt{\sigma_f^2 T}} \varphi(y) + o\left(\frac{1}{\sqrt{T}}\right)$$

uniformément en y .

C'est le résultat le plus général que nous connaissions sur l'asymptotique de la mesure empirique pour les processus de diffusion. Une borne de type Berry-Esseen est obtenue par Lezaud [60] dans le cadre des processus de Markov possédant un trou spectral et un semi-groupe réversible, pour des fonctions f bornées. Par ailleurs, un développement similaire est obtenu par Kontoyiannis et Meyn [50] dans le cadre général des processus de Markov géométriquement ergodiques, mais toujours pour une fonction f bornée. Enfin, mentionnons le travail récent de Bertail et Cléménçon [12] où apparaît un résultat analogue au notre dans le cadre des chaînes de Markov récurrentes et possédant un atome. Ce dernier article propose une ouverture vers la méthode du bootstrap. Nous reviendrons sur ce point à la fin du chapitre.

Remarquons la dépendance par rapport à la condition initiale qui apparaît uniquement au travers de la fonction $G_f(x)$. Ce terme reflète le biais de la moyenne empirique quand la condition initiale ne suit pas la mesure invariante.

4.2. Théorème de Représentation. La proposition 1.2.2 établit une convergence polynomiale du semi-groupe $p_f(t, x)$. Pour établir le théorème 4.1.1 nous avons besoin d'un résultat similaire pour le gradient $\nabla_x p_f(t, x)$ de cette fonction. Nous ne sommes pas parvenu à obtenir ce résultat sous les seules hypothèses (H1)–(H4) et requérons donc une plus grande régularité sur notre modèle en introduisant les hypothèses (H5) et (H6).

PROPOSITION 4.2.1. *Nous nous plaçons sous les hypothèses (H1)–(H6) et nous considérons*

$$p_f(t, x) = \mathbb{E}_x \left[\tilde{f}(X_t) \right].$$

Cette fonction est une solution de classe C^∞ de l'équation aux dérivées partielles

$$\begin{cases} \frac{d}{dt} p_f(t, x) &= \mathcal{L} p_f(t, x) \\ p_f(0, x) &= \tilde{f}(x). \end{cases}$$

Par conséquent, nous avons le théorème de représentation suivant

$$\tilde{f}(X_T) = \mathbb{E}_x \left[\tilde{f}(X_T) \right] + \int_0^T \sigma^*(X_t) \nabla_x p_f(T-t, X_t) \cdot dW_t.$$

De plus, pour tout $k > 0$, il existe des constantes strictement positives C_1 et s_1 telles que

$$|\nabla_x p_f(t, x)| \leq \frac{C_1(1 + |x|^{s_1})}{(1+t)^{k+1}}.$$

DÉMONSTRATION. Sous nos hypothèses, l'existence et la régularité de la solution du problème parabolique sont établis par Ladyženskaja, Solonnikov et Ural'ceva [57] chapitre 4. Le théorème de représentation est alors une simple conséquence de la formule d'Itô appliquée au processus $p_f(0, X_T)$.

Pour obtenir le contrôle du gradient, nous faisons appel aux arguments développés par D. Talay [83]. Cet article se place dans la cadre où $\alpha = 0$ (voir notre hypothèse (H2)). Il est clair que la preuve s'étend sans modification au cas $\alpha > 0$ et nous obtenons dans ce cas une convergence exponentielle. Nous allons maintenant vérifier que, dans le cas $-1 < \alpha < 0$, une légère modification de la preuve permet d'obtenir une convergence polynomiale. Notons tout d'abord que nos nouvelles hypothèses impliquent l'existence d'une densité p régulière pour la mesure μ .

Remarque préliminaire

Un résultat classique (voir Kunita [52]) nous assure que la fonction $p_f(t, x)$ vérifie

$$\forall m \in \mathbb{N}, \exists s_m \in \mathbb{N}, \forall t > 0, \exists C_m(t) > 0 : |D^m p_f(\theta, x)| \leq C_m(t)(1 + |x|^{s_m}), \forall \theta \leq t.$$

Ainsi, la fonction f , la fonction $p_f(t, x)$ et toutes ses dérivées appartiennent à $L^2(\mathbb{R}^n, \mu)$.

Première étape : contrôle de la fonction caractéristique loin de l'origine

Dans un premier temps, nous voulons établir que, pour toute boule B , il existe une constante $C > 0$ telle que

$$(77) \quad \forall t > 0, \forall x \in B, |p_f(t, x)| \leq \frac{C}{(1+t)^k}, \forall k \in \mathbb{N},$$

puis d'étendre cette majoration aux dérivées spatiales de la fonction $p_f(t, x)$. Pour ce faire, l'idée est de prouver que, d'une part,

$$\int |p_f(t, x)|^2 p(x) dx \leq \frac{C}{(1+t)^k}, \forall k \in \mathbb{N},$$

que, d'autre part, pour tout multi-indice J ,

$$(78) \quad \int |\partial_J p_f(t, x)|^2 p(x) dx \leq \frac{C_J}{(1+t)^k}, \forall k \in \mathbb{N},$$

puis d'appliquer le théorème d'injection de Sobolev pour conclure. La première assertion est une conséquence de la proposition 1.2.2. La seconde assertion se démontre par récurrence. Nous en détaillons la première étape. Il s'agit de prouver que

$$(79) \quad \forall h > 0, \int_0^{+\infty} \left(\int |\nabla p_f(t, x)|^2 d\mu \right) (1+t)^h dt < +\infty.$$

Or, avec la convention de sommation sur les indices, nous avons

$$\frac{d}{dt} |p_f(t, x)|^2 - \mathcal{L}(|p_f(t, x)|^2) = -a_{ij}(x)(\partial_i p_f(t, x))(\partial_j p_f(t, x))$$

et donc, en multipliant par $(1+t)^h$ et en intégrant par rapport à μ ,

$$(1+t)^h \frac{d}{dt} \int |p_f(t, x)|^2 d\mu + C(1+t)^h \int |\nabla p_f(t, x)|^2 d\mu \leq 0.$$

Nous intégrons maintenant cette inégalité entre 0 et un T suffisamment grand,

$$(1+T)^h \int |p_f(T, x)|^2 d\mu \leq \int |\tilde{f}|^2 d\mu \\ + C \int_0^T \left(\int |\nabla p_f(t, x)|^2 d\mu \right) (1+t)^h dt + h \int_0^T \left(\int |p_f(t, x)|^2 d\mu \right) (1+t)^{h-1} dt.$$

Nous utilisons maintenant la borne de la proposition 1.2.2 pour obtenir (79), pour $h \leq k$. Maintenant, il est possible de choisir des constantes C_1 et C_2 , strictement positives et telles que

$$\begin{aligned} \frac{d}{dt} |\nabla p_f(t, x)|^2 - \mathcal{L}(|\nabla p_f(t, x)|^2) &= -a_{ij}(\partial_{ik} p_f(t, x))(\partial_{jk} p_f(t, x)) \\ &\quad + (\partial_k a_{ij})(\partial_{ij} p_f(t, x))(\partial_k p_f(t, x)) \\ &\quad + 2(\partial_k b_i)(\partial_i p_f(t, x))(\partial_k p_f(t, x)) \\ &\leq -C_1 |D^2 p_f(t, x)|^2 + C_2 |\nabla p_f(t, x)|^2. \end{aligned}$$

En choisissant $k < h$ et en procédant comme précédemment, nous obtenons

$$(1+T)^k \int |\nabla p_f(T, x)|^2 d\mu \leq \int |\nabla \tilde{f}|^2 d\mu \\ + C_1 \int_0^T \left(\int |D^2 p_f(t, x)|^2 d\mu \right) (1+t)^k dt + (C_2 + k) \int_0^T \left(\int |\nabla p_f(t, x)|^2 d\mu \right) (1+t)^k dt.$$

Et nous démontrons ainsi que

$$(80) \quad \int |\nabla p_f(t, x)|^2 p(x) dx \leq \frac{C}{(1+t)^k}, \forall k \in \mathbb{N}.$$

Pour itérer ce raisonnement, nous supposons que, pour tout $m \leq l$, il existe des constantes $C_m > 0$ telles que

$$\int |D^m p_f(t, x)|^2 p(x) dx \leq \frac{C_m}{(1+t)^k}, \forall k \in \mathbb{N}.$$

Comme

$$|D^m p_f(t, x)|^2 = \sum_{J, |J|=l} (\partial_J p_f(t, x))^2,$$

pour tout multi indice J de taille $l+1$, il existe une famille F_J de multi indices de taille au plus $l+1$ et une famille de fonctions bornées $(\varphi_{KL}^J, K, L \in F_J)$ telles que

$$\begin{aligned} \frac{d}{dt} D_J p_f(t, x) &= D_J \mathcal{L} p_f(t, x) \\ &= b_i \partial_{J \cup \{i\}} p_f(t, x) + \frac{1}{2} a_{ij} \partial_{J \cup \{ij\}} p_f(t, x) \\ &\quad + \sum_{K, L, |K|+|L| \leq 2l_1} \varphi_{KL}^J \partial_K p_f(t, x) \partial_L p_f(t, x). \end{aligned}$$

Ainsi,

$$\begin{aligned}
\frac{d}{dt}|D^l p_f(t, x)|^2 - \mathcal{L}|D^l p_f(t, x)|^2 &= -a_{ij}\partial_{J\cup\{i\}}p_f(t, x)\partial_{J\cup\{j\}}p_f(t, x) \\
&+ \sum_{K, L, |K|+|L|\leq 2l+1} \varphi_{KL}^J \partial_K p_f(t, x)\partial_L p_f(t, x) \\
&\leq -C_1^l |D^{l+1} p_f(t, x)|^2 + C_2^l \sum_{m\leq l} |D^m p_f(t, x)|^2.
\end{aligned}$$

En procédant comme précédemment, nous choisissons une constante k suffisamment petite, nous multiplions l'inégalité précédente par $(1+t)^{h_{m+1}}$ et nous intégrons par rapport à μ de telle sorte que nous vérifions

$$\int_0^{+\infty} \left(\int |D^{l+1} p_f(t, x)|^2 d\mu \right) (1+t)^{k_{l+1}} dt < +\infty.$$

Puis nous écrivons

$$\frac{d}{dt}|D^{l+1} p_f(t, x)|^2 - \mathcal{L}|D^{l+1} p_f(t, x)|^2 \leq -C_1^{l+1}|D^{l+2} p_f(t, x)|^2 + C_2^{l+1} \sum_{m\leq l+1} |D^m p_f(t, x)|^2,$$

nous choisissons $h_{l+1} < k_{l+1}$ et nous procédons comme précédemment. Finalement, comme la densité p est une fonction strictement positive sur toute boule B sous nos nouvelles hypothèses, nous avons

$$\|\partial_J p_f(t, \cdot)\|_{L^2(B)}^2 \leq C \int |\partial_J p_f(t, x)|^2 p(x) dx,$$

et nous pouvons invoquer le théorème d'injection de Sobolev pour vérifier (77).

Deuxième étape : contrôle de la fonction caractéristique autour de l'origine

Nous définissons maintenant

$$\pi_s(x) := \frac{1}{(1+|x|^{\alpha+1})^s}.$$

La deuxième étape de la preuve consiste en la démonstration de la majoration

$$(81) \quad \int |p_f(t, x)|^2 \pi_s(x) dx \leq \frac{C}{(1+t)^k}, \quad \forall k \in \mathbb{N},$$

pour un $s \in \mathbb{N}$. Ce résultat est une conséquence directe de la proposition 1.2.2.

Conclusion

La troisième et dernière étape de la preuve réside en l'extension de la majoration précédente aux dérivées spatiales de $p_f(t, x)$. La démonstration est tout-à-fait similaire à celle de la première étape, en intégrant, non plus par rapport à la densité $p(x)$, mais contre la fonction $\frac{1}{(1+|x|^{\alpha+1})^s}$ pour une valeur de s suffisamment grande. Nous raisonnons donc une nouvelle fois par récurrence. Pour le multi indice I , nous définissons l'entier M_I par $|I| = [M_I - n/2]$ où

n est la dimension du processus X . Tout d'abord, nous savons par [83] qu'il existe un entier s_0 tel que

$$\forall t > 0, \forall s \geq s_0, \forall m \leq M_I, D^m p_f(t, x) \pi_s(x) \in L^2(\mathbb{R}^n).$$

Supposons que nous ayons prouvé

$$\exists C_I > 0 : \forall m \leq M_I, \forall t > 0, \int |D^m p_f(t, x)|^2 \pi_s(x) dx \leq \frac{C_I}{(1+t)^k}, \forall k \in \mathbb{N}.$$

Il est facile de vérifier que

$$(82) \quad \partial \pi_s(x) = \psi_J(x) \pi_s(x), \psi_J \text{ bornée.}$$

Nous aurions ainsi

$$\exists C_I > 0 : \forall m \leq M_I, \forall t > 0, \int |D^m p_f(t, x) \pi_s(x)|^2 dx \leq \frac{C_I}{(1+t)^k}, \forall k \in \mathbb{N}$$

et notre résultat se déduirait une nouvelle fois du théorème d'injection de Sobolev. Or, en utilisant une nouvelle fois la remarque (82), nous vérifions qu'il est suffisant de montrer que, pour tout multi indice J tel que $|J| \leq M_I$, il existe une constante $C_J > 0$ telle que

$$\int |\partial_J p_f(t, x)|^2 \pi_s(x) dx \leq \frac{C_J}{(1+t)^k}, \forall k \in \mathbb{N}.$$

Si $|J| = 1$, tout comme dans Talay [83], nous remarquons qu'il existe des fonctions φ_1 et φ_2 telles que

- φ_1 est une fonction bornée, indépendante de s ,
- φ_2 dépend de s mais converge vers 0 quand $|x| \rightarrow +\infty$,
- l'égalité suivante est vérifiée

$$\begin{aligned} \int \mathcal{L} |p_f(t, x)|^2 \pi_s(x) dx &= \int |p_f(t, x)|^2 \mathcal{L}^* \pi_s(x) dx \\ &= \int \left(\varphi_1(x) + \varphi_2(x) + 2s \frac{b(x) \cdot x}{1 + |x|^{\alpha+1}} \right) |p_f(t, x)|^2 \pi_s(x) dx. \end{aligned}$$

Il est alors possible, en choisissant une valeur de s suffisamment grande de trouver une boule B telle que

$$\forall x \in \mathbb{R}^n - B, \varphi_1(x) + \varphi_2(x) + 2s \frac{b(x) \cdot x}{1 + |x|^{\alpha+1}} < 0.$$

En faisant appel à (77), nous en déduisons l'existence d'une constante $C_0 > 0$ telle que

$$\int \mathcal{L} |p_f(t, x)|^2 \pi_s(x) dx \leq \frac{C_0}{(1+t)^k}, \forall k \in \mathbb{N}.$$

Il est alors possible de démontrer que les inégalités (79) et (80) restent valides en substituant $\pi_s(x)$ à $p(x)$ puis de généraliser aux dérivées d'ordre supérieur en procédant par récurrence. \square

4.3. Preuve du théorème 4.1.1.

DÉMONSTRATION. Ce théorème est une conséquence du théorème 2.1.1. En effet, d'après la représentation (65),

$$\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T \tilde{f}(X_s) ds = \frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s + \frac{1}{\sqrt{\sigma_f^2 T}} G_f(x) - \frac{1}{\sqrt{\sigma_f^2 T}} G_f(X_T).$$

La difficulté réside dans le fait que notre terme de reste $\frac{1}{\sqrt{\sigma_f^2 T}} G_f(X_T)$ est d'ordre $O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$ et non plus $o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$ comme c'était le cas dans l'étude des martingales normalisées. Par conséquent, nous ne pouvons appliquer la méthode du delta. Nous allons donc reprendre l'argument de Feller en considérant que nous n'avons pas de terme de reste et expliciter complètement les calculs.

Première étape : contrôle de la fonction caractéristique loin de l'origine

Grâce à (H7) nous pouvons procéder comme dans la première partie de la preuve du théorème 2.1.1.

Seconde étape : contrôle de la fonction caractéristique autour de l'origine

Tout comme pour le développement d'Edgeworth pour les martingales normalisées, les constantes du développement s'obtiennent à partir du développement des moments. Il est facile d'établir que

$$\mathbb{E} \left[\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T \tilde{f}(X_s) ds \right] = \frac{1}{\sqrt{\sigma_f^2 T}} G_f(x) + O\left(\frac{1}{T}\right),$$

et que

$$\text{Var} \left[\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T \tilde{f}(X_s) ds \right] = \text{Var} \left[\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s \right] + O\left(\frac{1}{T}\right).$$

Par contre, le développement du moment d'ordre 3 est plus difficile à obtenir.

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{\int_0^T \tilde{f}(X_s) ds - G_f(x)}{\sqrt{\sigma_f^2 T}} \right)^3 \right] &= \mathbb{E} \left[\left(\frac{\int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s - G_f(X_T)}{\sqrt{\sigma_f^2 T}} \right)^3 \right] \\
&= \mathbb{E} \left[\left(\frac{\int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s}{\sqrt{\sigma_f^2 T}} \right)^3 \right] - \mathbb{E} \left[\left(\frac{G_f(X_T)}{\sqrt{\sigma_f^2 T}} \right)^3 \right] \\
&\quad + 3 \mathbb{E} \left[\frac{\left(\int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s \right)^2 G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] \\
&\quad + 3 \mathbb{E} \left[\frac{\int_0^T (\sigma^* \nabla G_f)(X_s) \cdot dW_s G_f(X_T)^2}{\sqrt{\sigma_f^2 T}^3} \right].
\end{aligned}$$

Nous connaissons déjà le développement du premier terme grâce à l'équation (73). Le deuxième terme est clairement d'ordre $O\left(T^{-\frac{3}{2}}\right)$ et le dernier terme, d'ordre $O\left(T^{-1}\right)$. Nous renvoyons à la preuve de l'équation (73) pour vérifier cette dernière affirmation. La difficulté réside donc dans le troisième terme qui s'exprime également

$$\begin{aligned}
\mathbb{E} \left[\frac{\left(M_T^{\sigma^* \nabla G_f} \right)^2 G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] &= \mathbb{E} \left[\frac{\left\{ \left(M_T^{\sigma^* \nabla G_f} \right)^2 - \langle M^{\sigma^* \nabla G_f} \rangle_T \right\} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] \\
&\quad + \mathbb{E} \left[\frac{\langle M^{\sigma^* \nabla G_f} \rangle_T G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] \\
&= 2 \mathbb{E} \left[\frac{\int_0^T M_s^{\sigma^* \nabla G_f} dM_s^{\sigma^* \nabla G_f} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] + O\left(\frac{1}{T}\right).
\end{aligned}$$

Pour vérifier cette dernière équation, nous appliquons une nouvelle fois la représentation de Poisson au terme $\langle M^{\sigma^* \nabla G_f} \rangle_T$ puis nous raisonnons comme lors de la démonstration du développement (73). Nous avons un dernier terme pour lequel nous faisons appel au théorème de représentation de la proposition 4.2.1.

$$\mathbb{E} \left[\frac{\int_0^T M_s^{\sigma^* \nabla G_f} dM_s^{\sigma^* \nabla G_f} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] = \frac{\int_0^T \mathbb{E} \left[M_s^{\sigma^* \nabla G_f} (\sigma^*(X_s) \nabla G_f(X_s) \cdot \sigma^*(X_s) \nabla_x p_{G_f}(T-s, X_s)) \right] ds}{\sqrt{\sigma_f^2 T}^3}.$$

Nous utilisons maintenant les résultats de la proposition 4.2.1 et l'inégalité de Cauchy-Schwarz pour obtenir la majoration

$$\left| \mathbb{E} \left[\frac{\int_0^T M_s^{\sigma^* \nabla G_f} dM_s^{\sigma^* \nabla G_f} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] \right| \leq C \frac{\int_0^T \sqrt{\mathbb{E} [\langle M^{\sigma^* \nabla G_f} \rangle_s]} \frac{1+|x|^s}{(1+(T-s))^k} ds}{\sqrt{\sigma_f^2 T}^3}$$

pour tout $k \in \mathbb{N}$. Par ergodicité du processus X , nous avons également

$$\left| \mathbb{E} \left[\frac{\int_0^T M_s^{\sigma^* \nabla G_f} dM_s^{\sigma^* \nabla G_f} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] \right| \leq \frac{C(1+|x|^s)}{T} \int_0^T \frac{1}{(1+(T-s))^k} ds.$$

Il est alors immédiat à partir de cette expression de vérifier que

$$\mathbb{E} \left[\frac{\int_0^T M_s^{\sigma^* \nabla G_f} dM_s^{\sigma^* \nabla G_f} G_f(X_T)}{\sqrt{\sigma_f^2 T}^3} \right] = O\left(\frac{1}{T}\right).$$

Ces développements permettent d'adapter la seconde partie de la preuve du théorème 2.1.1. \square

4.4. Application en simulation. L'application la plus évidente de ce résultat en simulation est la choix d'une durée de simulation T pour obtenir une précision donnée ε . Il est possible d'estimer grossièrement une telle durée à partir du TCL (66), mais nous avons maintenant les moyens d'être beaucoup plus précis. En effet, admettons que nous cherchions à obtenir une précision ε avec probabilité α . Nous recherchons donc une valeur de T telle que

$$\mathbb{P} \left(\left| \frac{1}{T} \int_0^T \tilde{f}(X_s) ds \right| \leq \varepsilon \right) \geq 1 - \alpha.$$

Un intervalle de confiance symétrique corrigé pour la moyenne recherchée – qui tient en particulier compte du biais de l'estimateur – est

$$\left[\frac{1}{T} \int_0^T f(X_s) ds + \left(\frac{\mu(\sigma^* \nabla G_f \cdot \sigma^* \nabla G_{|\sigma^* \nabla G_f|^2}) h_2(q_\alpha)}{2\sigma_f^2 T} + \frac{G_f(x)}{\sigma_f T} \right) \pm \frac{\sigma_f q_\alpha}{\sqrt{T}} \right]$$

avec q_α , le quantile de niveau α de la loi gaussienne standard. Ce résultat s'obtient en remarquant qu'un développement de Taylor dans le premier terme du développement d'Edgeworth du théorème 4.1.1 conduit à

$$\mathbb{P}_x \left(\frac{1}{\sqrt{\sigma_f^2 T}} \int_0^T \tilde{f}(X_s) ds + \left(\frac{\mu(\sigma^* \nabla G_f \cdot \sigma^* \nabla G_{|\sigma^* \nabla G_f|^2}) h_2(y)}{2\sigma_f^3 \sqrt{T}} + \frac{G_f(x)}{\sqrt{\sigma_f^2 T}} \right) \leq y \right) = \Phi(y) + o\left(\frac{1}{\sqrt{T}}\right)$$

uniformément en y . Pour parvenir à une simulation de la précision ε désirée, il faudrait donc choisir une durée de simulation T tel que

$$T \geq \frac{1}{\varepsilon^2} \left(\sigma_f q_\alpha - \sqrt{\sigma_f^2 q_\alpha^2 + 4\varepsilon \left(\frac{\mu(\sigma^* \nabla G_f \cdot \sigma^* \nabla G_{|\sigma^* \nabla G_f|^2}) h_2(q_\alpha)}{2\sigma_f^2 T} + \frac{G_f(x)}{\sigma_f T} \right)} \right)^2.$$

Nous obtenons bien une précision de la valeur déduite du TCL (66). Bien entendu, ce résultat théorique requière en pratique l'estimation des quantités $\mu (\sigma^* \nabla G_f \cdot \sigma^* \nabla G_{|\sigma^* \nabla G_f|^2})$, $G_f(x)$ et σ_f^2 . Or, nous ne disposons d'un estimateur simple que pour cette dernière variable (voir l'équation 56 page 32 du chapitre 2). Néanmoins, il est certainement possible d'obtenir des estimations a priori sur certains exemples particuliers.

5. Développements de Cornish-Fisher

Pour terminer ce chapitre, nous nous intéressons à une conséquence particulièrement intéressante des développements d'Edgeworth obtenus précédemment : le développement asymptotique des quantiles de variables aléatoires ou développement de Cornish-Fisher. Le résultat général est le suivant.

5.1. Résultat principal. Le résultat suivant est un corrolaire classique des développements de type Edgeworth.

THÉORÈME 5.1.1. *Soit une variable aléatoire S_T admettant un développement d'Edgeworth de la forme*

$$\mathbb{P}(S_T \leq y) = \Phi(y) - \frac{1}{\sqrt{T}} h(y) \varphi(y) + o\left(\frac{1}{\sqrt{T}}\right).$$

et soit

$$\omega_\alpha := \inf_{y \in \mathbb{R}} \{\mathbb{P}(S_T \leq y) \geq \alpha\}.$$

Alors

$$\omega_\alpha = q_\alpha + \frac{1}{\sqrt{T}} h(q_\alpha) + o\left(\frac{1}{\sqrt{T}}\right)$$

uniformément en $\varepsilon < \alpha < 1 - \varepsilon$ pour tout $\varepsilon \in]0, \frac{1}{2}[$ avec q_α , le quantile de niveau α de la loi gaussienne centrée réduite.

Il est clair que ce résultat s'applique aux trois théorèmes démontrés précédemment.

DÉMONSTRATION. La preuve est très classique. Elle repose sur un développement de Taylor dans le premier terme du développement d'Edgeworth au voisinage du point q_α . Nous renvoyons par exemple à Hall [41] pour les détails. \square

Si les coefficients du polynôme h étaient explicitement connus – ou simples à estimer – l'erreur de couverture d'un intervalle de confiance de niveau α pourrait aisément être corrigée en utilisant le fait que

$$\mathbb{P}\left(S_T \leq y - \frac{1}{\sqrt{T}} h(y)\right) = \Phi(y) + o\left(\frac{1}{\sqrt{T}}\right),$$

uniformément pour tout $y \in I$, intervalle borné de \mathbb{R} . Evidemment, dans notre cas, ces coefficients sont très difficiles à estimer. Le développement de Cornish-Fisher permet néanmoins de corriger l'erreur de couverture grâce à une méthode de bootstrap. C'est l'objet de la section suivante.

5.2. Application au bootstrap des intervalles de confiance. Le bootstrap appartient à la famille des techniques statistiques modernes qui exploitent la méthode de Monte-Carlo dans le but d'obtenir des estimateurs précis et des tests statistiques puissants pour les modèles complexes. L'inférence indirecte de Gourieroux, Monfort et Renault [40] et les méthodes de Monte-Carlo par chaînes de Markov (MCMC) sont deux autres membres bien connus de cette famille.

Le principe du bootstrap (de l'expression *Pull yourself up by your own bootstraps*), bien que présent depuis longtemps sous différentes formes, a été formalisé en 1979 par B. Efron [27]. Soient $(X_i)_{i=1..n}$, un échantillon de loi \mathbb{P} et $H_n := H_n(X_1, \dots, X_n; \mathbb{P})$, une statistique dont nous cherchons à construire un estimateur. Si \mathbb{Q}_n converge en probabilités vers \mathbb{P} et que $(X_i^*)_{i=1..n}$ dénote un échantillon de loi \mathbb{Q}_n , il est raisonnable de penser que, sous des hypothèses appropriées, la loi de $H_n^* := H_n(X_1^*, \dots, X_n^*; \mathbb{Q}_n)$ va approcher celle de H_n . L'idée d'Efron est de choisir $\mathbb{Q}_n := \mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, la distribution empirique des variables X_i . Dans la mesure où $\mathbb{P}_n \rightarrow \mathbb{P}$ presque sûrement et où simuler des échantillons de loi \mathbb{P}_n est très simple, cette méthode très naturelle connut un grand succès.

Depuis son origine, elle s'est étendue à pratiquement toutes les questions d'estimation du cadre discret. Toutefois, à notre connaissance, il n'existe pas d'étude sur l'efficacité du bootstrap appliqué à la statistique des processus de diffusion. Cette question a toutefois été abordée d'un point de vue empirique par Conley et Hansen [24] et Phillips et Yu [74].

Par ailleurs, il est connu, entre autre du fait de l'ouvrage de P. Hall [41], que les développements de Cornish-Fisher sont le fondement théorique du bootstrap des intervalles de confiance. Nous nous proposons donc pour terminer ce chapitre d'illustrer ce principe dans le cadre des processus de diffusion. Et en particulier sur l'exemple de l'estimation du paramètre de dérive par maximum de vraisemblance, dans le cas unidimensionnel. Nous nous plaçons donc dans le cadre où

$$S_T = \sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)} T \left(\hat{\theta}_T - \theta \right).$$

Un intervalle de confiance asymptotique de niveau α pour le paramètre θ peut être

$$I_T(\alpha) = \left[\hat{\theta}_T - \frac{q_\alpha}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)} T}, +\infty \right].$$

D'après la proposition 3.2.1, l'erreur de couverture de cet intervalle de confiance est

$$\mathbb{P}(\theta \in I_T(\alpha)) - \alpha = -\frac{h_\theta(q_\alpha)}{\sqrt{T}} + o\left(\frac{1}{\sqrt{T}}\right)$$

avec

$$h_\theta(q_\alpha) = \frac{\varphi(q_\alpha)}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)}} \left(\frac{\mu \left(bG'_{\frac{b^2}{\sigma^2}} \right) h_2(q_\alpha)}{2\mu \left(\frac{b^2}{\sigma^2} \right)} \right) + \frac{\varphi(q_\alpha)}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)}} \frac{\mu \left(bG'_{\frac{b^2}{\sigma^2}} \right)}{\mu \left(\frac{b^2}{\sigma^2} \right)}.$$

Le polynôme h_θ dépend évidemment de la vraie valeur du paramètre θ et ce, au travers de la mesure invariante μ et de la fonction $G_{\frac{b^2}{\sigma^2}}$. Il est clair que si le quantile ω_α était connu, cette

erreur serait nulle. L'idée du bootstrap est donc d'estimer ce quantile en supposant que $\hat{\theta}_T$ est la vraie valeur du paramètre θ . Nous définissons donc

$$\tilde{\omega}_\alpha(T) := \inf_{y \in \mathbb{R}} \left\{ \mathbb{P} \left(\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)} T \left(\hat{\theta}_T - \hat{\theta}_T \right) \leq y \right) \geq \alpha \right\}$$

où $\hat{\theta}_T$ est construit à partir de l'observation du processus simulé

$$dY_t = \hat{\theta}_T b(Y_t) dt + \sigma(X_t) dB_t, \quad t \leq T, \quad Y_0 = X_0$$

sur le modèle de $\hat{\theta}_T$:

$$\hat{\theta}_T = \frac{\int_0^T \frac{b(Y_t)}{\sigma^2(Y_t)} dY_t}{\int_0^T \frac{b^2(Y_t)}{\sigma^2(Y_t)} dt}$$

avec B , un mouvement brownien indépendant de W . Le quantile $\tilde{\omega}_\alpha(T)$ peut donc être estimé par la méthode de Monte-Carlo. Nous définissons maintenant l'intervalle de confiance bootstrap

$$I_T^{\text{boot}}(\alpha) = \left[\hat{\theta}_T - \frac{\tilde{\omega}_\alpha(T)}{\sqrt{\mu \left(\frac{b^2}{\sigma^2} \right)} T}, +\infty \right[.$$

Or, conditionnellement à $\hat{\theta}_T$, le quantile $\tilde{\omega}_\alpha$ admet le développement

$$\tilde{\omega}_\alpha(T) = q_\alpha + \frac{1}{\sqrt{T}} h_{\hat{\theta}_T}(q_\alpha) + o\left(\frac{1}{\sqrt{T}}\right)$$

tout comme le vrai quantile se développe en

$$\omega_\alpha = q_\alpha + \frac{1}{\sqrt{T}} h_\theta(q_\alpha) + o\left(\frac{1}{\sqrt{T}}\right).$$

Par conséquent, nous avons la relation

$$\tilde{\omega}_\alpha(T) = \omega_\alpha + \frac{1}{\sqrt{T}} (h_{\hat{\theta}_T}(q_\alpha) - h_\theta(q_\alpha)) + o\left(\frac{1}{\sqrt{T}}\right).$$

Si la fonction $\theta \mapsto h_\theta$ est continue, alors, comme $\hat{\theta}_T$ est un estimateur fortement consistant de θ , nous avons

$$h_{\hat{\theta}_T}(q_\alpha) - h_\theta(q_\alpha) \xrightarrow[T \rightarrow +\infty]{\text{p.s.}} 0$$

et

$$\tilde{\omega}_\alpha(T) = \omega_\alpha + o\left(\frac{1}{\sqrt{T}}\right).$$

Il ne reste plus qu'à procéder à un développement de Taylor dans le premier terme du développement d'Edgeworth de la proposition 3.2.1 pour s'assurer que

$$\mathbb{P}(\theta \in I_T^{\text{boot}}(\alpha)) - \alpha = o\left(\frac{1}{\sqrt{T}}\right).$$

Nous avons donc corrigé l'erreur de couverture de l'intervalle de confiance.

REMARQUE. Pour vérifier que le gain de cette procédure numériquement coûteuse n'est pas négligeable, nous aurions besoin d'un développement d'ordre supérieur. De même, pour corriger l'erreur de couverture des intervalles de confiance symétriques (qui est déjà d'ordre $o\left(\frac{1}{\sqrt{T}}\right)$) une plus grande précision est nécessaire. Il est légitime de supposer que le terme suivant dans un tel développement serait d'ordre $O\left(\frac{1}{T}\right)$ bien que nous n'ayons pas établi ce résultat.

6. Conclusion

En nous fondant sur les propriétés des martingales, nous avons donc pu établir des développements de type Edgeworth pour différentes quantités d'intérêt en statistique des processus. Ces résultats ne sont pour la plupart pas complètement nouveaux. Toutefois, notre approche nous a permis d'affaiblir les hypothèses supposées dans les travaux antérieurs. De plus, comme elle ne repose sur aucune propriété fondamentale de l'espace de Wiener – contrairement aux stratégies fondées sur le calcul de Malliavin – nous pensons que notre technique de preuve pourrait se généraliser et permettre d'obtenir des résultats similaires pour des fonctionnelles additives martingales de processus de Markov ergodiques plus généraux. Une autre question d'intérêt serait l'obtention d'un développement d'Edgeworth pour la moyenne empirique studentisée par l'estimateur de la variance issu du TCL presque sûr (voir le chapitre 2). Ce problème est abordé dans le cadre de chaînes de Markov possédant un atome par Bertail et Cléménçon [12].

STATISTICAL ESTIMATION OF SKEW PROCESSES

En collaboration avec Miguel Martinez (INRIA)

1. Introduction

The skew Brownian motion arises in the modelling of discontinuous physical systems. Geophysical phenomena, nuclear reactors, polymers, composite structures are examples of such systems. However, if the estimation of Feller diffusion processes has been extensively studied, so far, no result is available for the estimation of skewed processes. This article is a first attempt to calibrate their coefficients. First, let us make precise what we call skew processes.

Let $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space and let B be a \mathbb{F} -brownian motion. We consider a continuous stochastic process reflected over an interval $[a, b]$ and we set $a < \gamma < b$. This process X^σ is defined as the unique weak solution of the following stochastic differential equation :

$$(83) \quad X_t^\sigma = x + \int_0^t \sigma(X_s^\sigma) dB_s + (2\alpha - 1)L_t^\gamma(X^\sigma) - K_t^{[a,b]}(X^\sigma), \quad x \in [a, b], \quad \alpha \in]0, 1[$$

where $K^{[a,b]}(X^\sigma) := L^a(X^\sigma) - L^b(X^\sigma)$ and the process $L^\gamma(X^\sigma)$ is the symmetric local time of the process X^σ (that is the local time defined from the Tanaka formula with the convention $\text{sign}(0) = 0$; for references about local times, see Revuz and Yor [77]). The function σ is supposed to be Lipschitz and bounded away from zero. This function is supposed to be known.

The coefficient α is called the **skew coefficient** and the coefficient γ , the **locus**. Our main interest is in the estimation of these parameters from a single observation of the process X^σ in the long run. Thus, we shall consider the statistical model

$$(\Omega, \mathcal{F}, \mathbb{F}, (X_t^\sigma)_{0 \leq t \leq T}, C([a, b]), \mathcal{B}(C([a, b])), \mathbb{P}_{(\alpha, \gamma)}, \alpha \in]0, 1[, \gamma \in]a, b])$$

where $\mathbb{P}_{(\alpha, \gamma)}$ is the law of the process X^σ given α and γ .

As our estimation procedure relies on the ergodicity of the process X^σ , we introduce a reflection over a bounded interval by the mean of the process $K^{[a,b]}(X^\sigma)$. In this setting, M. Martinez proves the existence and uniqueness of a weak solution to equation (83) in [67]. This is sufficient for our purpose. Moreover, the hypothesis of a process living in a bounded domain seems suitable for practical purposes.

Figure 1 shows a typical path of the process X^σ when $\sigma \equiv 1$.

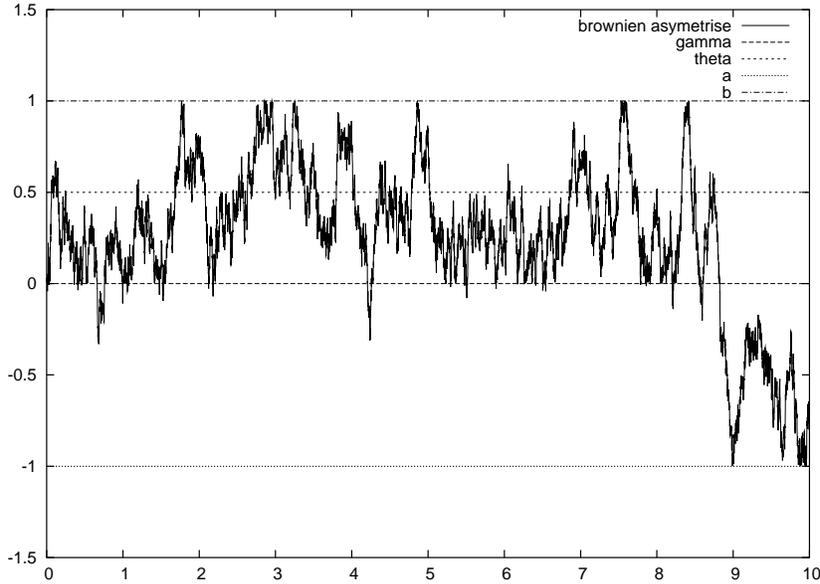


FIG. 1. A sample path of X^1 when $a = -1.0$, $b = 1.0$, $\alpha = 0.8$, $\gamma = 0.0$

Note that the coefficient σ can be estimated with the usual quadratic variation based estimators.

This paper is organized as follows. In section 2, we shall prove some results concerning the ergodicity of the process X^σ and construct an efficient estimator of the parameter α supposing that the locus γ is known. Then, in section 3, we built up a procedure to estimate adaptively the couple (α, γ) . Finally, some numerical results are presented to emphasize the accuracy of our method.

2. Estimation of α when γ is known

In this first part, we shall suppose that the locus of the asymetry γ is known.

2.1. Main Assumptions and Results. Set

$$(84) \quad M_T := \frac{1}{T} \int_0^T \sigma^2(X_s^\sigma) \mathbb{1}_{[\gamma, b]}(X_s^\sigma) ds, \quad T > 0,$$

and

$$(85) \quad \hat{\alpha}_T := \frac{M_T \int_a^\gamma \frac{dx}{\sigma^2(x)}}{M_T \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) + (b - \gamma)}, \quad T > 0.$$

Our main result is the following theorem.

THEOREM 2.1.1. *The random variable $\hat{\alpha}_T$ is a strongly consistent and asymptotically normal estimator of the coefficient α .*

This theorem will appear as a corollary of two propositions. The first one states an ergodic theorem for the process X^σ .

PROPOSITION 2.1.1. *Let ν^1 and ν^σ be the measures respectively defined by*

$$(86) \quad d\nu^1(x) := \frac{1}{\alpha(b - \gamma) + (1 - \alpha)(\gamma - a)} (\alpha \mathbb{1}_{[\gamma, b]}(x) + (1 - \alpha) \mathbb{1}_{[a, \gamma]}(x)) dx$$

and

$$(87) \quad d\nu^\sigma(x) := \frac{\frac{d\nu^1(x)}{\sigma^2(x)}}{\int_{\mathbb{R}} \frac{d\nu^1(x)}{\sigma^2(x)}}.$$

For any measurable function f in $L^1(\nu^\sigma)$,

$$(88) \quad \frac{1}{T} \int_0^T f(X_s^\sigma) ds \xrightarrow[T \rightarrow +\infty]{a.s.} \int f d\nu^\sigma.$$

REMARK. It is straightforward to check that the function σ^{-2} is in $L^1(\nu^1)$ from our hypotheses.

The second proposition of interest provides a central limit theorem.

PROPOSITION 2.1.2. *For any measurable function f in $L^1(\nu^\sigma)$,*

$$(89) \quad \sqrt{T} \left(\frac{1}{T} \int_0^T f(X_s^\sigma) ds - \int f d\nu^\sigma \right) \xrightarrow[T \rightarrow +\infty]{Law} \mathcal{N}(0, \sigma_f^2),$$

where $\mathcal{N}(0, \sigma_f^2)$ stands for the zero mean gaussian law with variance

$$(90) \quad \sigma_f^2 = 4 \int_{\mathbb{R}} \sigma(z)^2 \left(\int_z^b \frac{f(y) - \int f d\nu^\sigma}{\sigma(y)^2} dy \right)^2 d\nu^\sigma(z).$$

The next part of this paper is devoted to the proofs of these two propositions and of theorem 2.1.1.

2.2. Proofs. In order to prove the proposition 2.1.1, we start by considering the case $\sigma \equiv 1$. Then we shall deduce the general case using a time change argument.

LEMMA 2.2.1. If the diffusion function σ is constant and equal to 1, then the process X^1 has a unique invariant measure ν^1 defined by (86).

PROOF. If $\sigma \equiv 1$, equation (83) reads

$$(91) \quad X_t^1 = x + B_t + (2\alpha - 1)L_t^\gamma(X^1) - K_t^{[a, b]}(X^1), \quad x \in [a, b], \quad \alpha \in]0, 1[, \quad t \in [0, T].$$

Let $\varphi_{p, q}$ be the continuous function defined by

$$\varphi_{p, q}(y) := p(y - \gamma) \mathbb{1}_{(y < \gamma)} + q(y - \gamma) \mathbb{1}_{(y > \gamma)} + \gamma, \quad y \in \mathbb{R}; \quad p, q \in \mathbb{R}^*.$$

This function has an inverse $\varphi_{p, q}^{-1}$:

$$\varphi_{p, q}^{-1}(y) = \frac{1}{p}(y - \gamma) \mathbb{1}_{(y < \gamma)} + \frac{1}{q}(y - \gamma) \mathbb{1}_{(y > \gamma)} + \gamma, \quad y \in \mathbb{R}.$$

We consider the process $Y_t := \varphi_{p,q}(X_t^1)$. The Itô-Tanaka formula leads to

$$\begin{aligned}
\varphi_{p,q}(X_t^1) &= \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(X_s^1) dX_s^1 + \frac{q-p}{2} L_t^\gamma(X^1) \text{ with } \varphi'_{p,q}(y) := p\mathbb{1}_{(y<\gamma)} + q\mathbb{1}_{(y>\gamma)} \\
&= \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(X_s^1) dB_s + (2\alpha - 1) \int_0^t \varphi'_{p,q}(X_s^1) dL_s^\gamma(X^1) \\
&\quad - \int_0^t \varphi'_{p,q}(X_s^1) dK_s^{[a,b]}(X^1) + \frac{q-p}{2} L_t^\gamma(X^1) \\
&= \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(X_s^1) dB_s + \left[(2\alpha - 1) \frac{p+q}{2} + \frac{q-p}{2} \right] L_t^\gamma(X^1) \\
&\quad - \int_0^t \varphi'_{p,q}(X_s^1) dK_s^{[a,b]}(X^1).
\end{aligned}$$

We consider independently each integral in the former equation. First, note that $\varphi'_{p,q} \circ \varphi_{p,q}^{-1} \equiv \varphi'_{p,q}$ so that

$$\int_0^t \varphi'_{p,q}(X_s^1) dB_s = \int_0^t \varphi'_{p,q}(Y_s) dB_s.$$

So far, we have shown that

$$(92) \quad Y_t = \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(Y_s) dB_s + \left[(2\alpha - 1) \frac{p+q}{2} + \frac{q-p}{2} \right] L_t^\gamma(X^1) - K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y)$$

with $K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y) := \int_0^t \varphi'_{p,q}(X_s^1) dK_s^{[a,b]}(X^1)$. It is easy to check that this process is nothing else than

$$K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y) = L_t^{\varphi_{p,q}(a)}(Y) - L_t^{\varphi_{p,q}(b)}(Y).$$

Now, making use of the properties of the local time (see Revuz and Yor [77]), we have

$$\begin{aligned}
L_t^\gamma(X^1) &= L_t^\gamma(\varphi_{p,q}^{-1}(Y)) \text{ as } \varphi_{p,q}(\gamma) = \gamma \\
&= \frac{1}{2} D_- \varphi_{p,q}^{-1}(\gamma) L_t^{\gamma-}(Y) + \frac{1}{2} D_+ \varphi_{p,q}^{-1}(\gamma) L_t^{\gamma+}(Y)
\end{aligned}$$

where D_- (resp. D_+) stands for the left (resp. the right) derivative and $L_t^{\gamma-}(Y)$ (resp. $L_t^{\gamma+}(Y)$), for the left (resp. the right) local time of the process Y in γ (that is the local times defined from the Tanaka formula with, respectively, $\text{sign}(0) = -1$ and $\text{sign}(0) = 1$). Thus, we have

$$(93) \quad L_t^\gamma(X^1) = \frac{1}{2p} L_t^{\gamma-}(Y) + \frac{1}{2q} L_t^{\gamma+}(Y).$$

We want to rewrite this last expression in term of the symmetric local time of Y . In fact, we are looking for a constant β such that equation (92) can be written as

$$Y_t = \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(Y_s) dB_s + \beta L_t^\gamma(Y) - K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y).$$

We are going to make use of the two following relationships between the different notions of local time (see again Revuz and Yor [77]) :

$$(94) \quad \begin{cases} L_t^{\gamma^+}(Y) + L_t^{\gamma^-}(Y) &= 2L_t^\gamma(Y) \\ L_t^{\gamma^+}(Y) - L_t^{\gamma^-}(Y) &= \beta L_t^\gamma(Y) \end{cases}$$

This last equation is obtained from the general property

$$L_t^{\gamma^+}(Y) - L_t^{\gamma^-}(Y) = \int_0^t \mathbb{1}_{(Y_s=\gamma)} dV_s$$

where V is the finite variation part of the process Y . The process $K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y)$ only grows on the boundaries $\varphi_{p,q}(a)$ and $\varphi_{p,q}(b)$ and thus gives no contribution to the integral $\int_0^t \mathbb{1}_{(Y_s=\gamma)} dV_s$. From equation (94), we get

$$\begin{cases} L_t^{\gamma^+}(Y) &= \left(1 + \frac{\beta}{2}\right) L_t^\gamma(Y) \\ L_t^{\gamma^-}(Y) &= \left(1 - \frac{\beta}{2}\right) L_t^\gamma(Y) \end{cases}$$

Then, using equations (92) and (93), we see that β must satisfy

$$\beta = \left[(2\alpha - 1) \frac{p+q}{2} + \frac{q-p}{2} \right] \left[\frac{1}{2p} \left(1 - \frac{\beta}{2}\right) + \frac{1}{2q} \left(1 + \frac{\beta}{2}\right) \right].$$

So that, after simplifications, we have

$$(95) \quad \beta = \frac{2 \left[(2\alpha - 1) \frac{p+q}{2} + \frac{q-p}{2} \right] (p+q)}{4pq - \left[(2\alpha - 1) \frac{p+q}{2} + \frac{q-p}{2} \right] (p-q)}.$$

If we choose our constants p and q such that

$$\alpha = \frac{q}{p+q},$$

one can check that

$$\beta = \frac{q^2 - p^2}{q^2 + p^2}.$$

Therefore, the process Y satisfies the following SDE in the weak sense :

$$Y_t = \varphi_{p,q}(x) + \int_0^t \varphi'_{p,q}(Y_s) dB_s + \frac{q^2 - p^2}{q^2 + p^2} L_t^\gamma(Y) - K_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(Y).$$

It is known (see Martinez [67] for a recent proof) that such a process is associated to the divergence form operator

$$\mathcal{A} = \frac{d}{dx} \left(\frac{1}{2} \varphi'_{p,q} \frac{d}{dx} \right) + \text{boundary terms}$$

and therefore has a unique invariant measure, namely the uniform measure over the interval $[\varphi_{p,q}(a), \varphi_{p,q}(b)]$. Consequently, the process $X_t = \varphi_{p,q}^{-1}(Y_t)$ also has a unique invariant measure

whom we are now going to give the density $d\nu$. For any continuous and bounded function f , we must verify

$$\begin{aligned}\int f d\nu &= \frac{1}{\varphi_{p,q}(b) - \varphi_{p,q}(a)} \int_{\varphi_{p,q}(a)}^{\varphi_{p,q}(b)} f \circ \varphi_{p,q}^{-1}(y) dy \\ &= \frac{1}{\varphi_{p,q}(b) - \varphi_{p,q}(a)} \int_a^b f(y) \varphi'_{p,q}(y) dy.\end{aligned}$$

Therefore, as $\varphi_{p,q}(b) = q(b - \gamma) - \gamma$, $\varphi_{p,q}(a) = p(a - \gamma) - \gamma$ and $\varphi'_{p,q}(y) = p\mathbb{1}_{(y < \gamma)} + q\mathbb{1}_{(y > \gamma)}$, we have

$$\begin{aligned}d\nu(x) &= \frac{1}{q(b - \gamma) - p(a - \gamma)} (p\mathbb{1}_{[a, \gamma]}(x) + q\mathbb{1}_{[\gamma, b]}(x)) dx \\ &= \frac{1}{\alpha(b - \gamma) + (1 - \alpha)(\gamma - a)} ((1 - \alpha)\mathbb{1}_{[a, \gamma]}(x) + \alpha\mathbb{1}_{[\gamma, b]}(x)) dx \text{ as } \alpha = \frac{q}{p + q}.\end{aligned}$$

Thus, $\nu = \nu^1$ and the lemma is demonstrated. \square

The following corollary is an immediate consequence of the former lemma.

COROLLARY 2.2.1. If the diffusion function σ is constant and equal to 1, then, for any measurable function f in $L^1(\nu^1)$,

$$(96) \quad \frac{1}{T} \int_0^T f(X_s^1) ds \xrightarrow[T \rightarrow +\infty]{a.s.} \int f d\nu^1.$$

PROOF. The proof of this result is classical (see, e.g., Gihman and Skorohod [37]). The main assumption needed in the proof, the strong Markov property, holds for the process X^1 (see Martinez [67]). \square

PROOF (of proposition 2.1.1). We are now in position to demonstrate the proposition 2.1.1. We introduce the stopping time

$$\tau_t := \inf \left(s : \int_0^s \sigma^2(X_u^\sigma) du > t \right)$$

and the process $\tilde{Y}_t := X_{\tau_t}^\sigma$ which verifies

$$(97) \quad \tilde{Y}_t = x + \int_0^{\tau_t} \sigma(X_s^\sigma) dB_s + (2\alpha - 1)L_{\tau_t}^\gamma(X^\sigma) - K_{\tau_t}^{[a, b]}(X^\sigma).$$

The Dubins-Schwarz theorem states that there exists a Brownian motion W such that

$$\int_0^{\tau_t} \sigma(X_s^\sigma) dB_s = W_t, \forall t \geq 0, \text{ a.s.}$$

Moreover, a result of D. Revuz and M. Yor [77] asserts that $L_{\tau_t}^\gamma(X^\sigma) = L_t^\gamma(X_\tau^\sigma)$. Thus, equation (97) can be written as

$$\begin{aligned}\tilde{Y}_t &= x + W_t + (2\alpha - 1)L_t^\gamma(X_\tau^\sigma) - K_t^{[a, b]}(X_\tau^\sigma) \\ &= x + W_t + (2\alpha - 1)L_t^\gamma(\tilde{Y}) - K_t^{[a, b]}(\tilde{Y}).\end{aligned}$$

This is equation (91). It is known that such an equation has a unique weak solution. And we have yet demonstrated that this solution has a unique invariant measure. Therefore, lemma (2.2.1) applies. Let us recall the definition of the measure ν^σ :

$$d\nu^\sigma(x) = \frac{\frac{d\nu^1(x)}{\sigma^2(x)}}{\int_{\mathbb{R}} \frac{d\nu^1(x)}{\sigma^2(x)}},$$

and let us consider a measurable function f in $L^1(\nu^\sigma)$. We are interested in the convergence of the random variable

$$\frac{1}{\tau_T} \int_0^{\tau_T} f(X_s^\sigma) ds.$$

An easy change of variables allows to rewrite it

$$\frac{1}{\tau_T} \int_0^{\tau_T} f(X_s^\sigma) ds = \frac{\int_0^T f(\tilde{Y}_s) \frac{ds}{\sigma^2(\tilde{Y}_s)}}{\int_0^T \frac{ds}{\sigma^2(\tilde{Y}_s)}}.$$

Then, applying corollary 2.2.1, we can state that

$$\frac{1}{\tau_T} \int_0^{\tau_T} f(X_s^\sigma) ds \xrightarrow[T \rightarrow +\infty]{a.s.} \int f d\nu^\sigma.$$

As the process $(\tau_t)_{t \geq 0}$ is strictly increasing we can conclude the proof and assert that

$$\frac{1}{T} \int_0^T f(X_s^\sigma) ds \xrightarrow[T \rightarrow +\infty]{a.s.} \int f d\nu^\sigma.$$

□

So far, we have proved the strong law of large numbers. We now turn to the proof of the central limit theorem.

PROOF (of proposition 2.1.2). Once more, consider the function

$$\varphi_{p,q}(y) = p(y - \gamma) \mathbb{1}_{(y \leq \gamma)} + q(y - \gamma) \mathbb{1}_{(y \geq \gamma)} + \gamma, \quad y \in \mathbb{R}.$$

We now choose the parameters p and q such that

$$\alpha q + (\alpha - 1)p = 0.$$

Then, the process $\hat{Y}_t := \varphi_{p,q}(X_t^\sigma)$ as defined in (92) satisfies

$$(98) \quad d\hat{Y}_t = \rho(\hat{Y}_t) dB_t - dK_t^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(\hat{Y}), \quad t \in [0, T], \quad \hat{Y}_0 = \varphi_{p,q}(x),$$

where

$$\rho(y) := \sigma \circ \varphi_{p,q}^{-1}(y) \cdot (p \mathbb{1}_{(y \leq \gamma)} + q \mathbb{1}_{(y \geq \gamma)}).$$

Consider a measurable function f in $L^1(\nu^\sigma)$ and the Poisson problem

$$(99) \quad \begin{cases} \frac{1}{2} \rho(y)^2 \frac{d^2 u}{dy^2}(y) = -f \circ \varphi_{p,q}^{-1}(y) + \int f d\nu^\sigma, & y \in (\varphi_{p,q}(a), \varphi_{p,q}(b)) \\ \frac{du}{dy}(y) = 0, & y = \varphi_{p,q}(a) \text{ or } \varphi_{p,q}(b). \end{cases}$$

This problem has a solution twice differentiable, namely

$$u(x) := 2 \int_{\varphi_{p,q}(a)}^x \left(\int_z^{\varphi_{p,q}(b)} \frac{f \circ \varphi_{p,q}^{-1}(y) - \int f d\nu^\sigma}{\rho(y)^2} dy \right) dz, \quad x \in [\varphi_{p,q}(a), \varphi_{p,q}(b)].$$

By Itô's formula, we get

$$\begin{aligned}
u(\hat{Y}_T) &= u(\hat{Y}_0) + \frac{1}{2} \int_0^T \rho(\hat{Y}_s)^2 \frac{d^2 u}{dy^2}(\hat{Y}_s) ds + \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dB_s \\
&\quad - \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dK_s^{[\varphi_{p,q}(a), \varphi_{p,q}(b)]}(\hat{Y}) \\
&= u(\hat{Y}_0) - \int_0^T f \circ \varphi_{p,q}^{-1}(\hat{Y}_s) ds + T \int f d\nu^\sigma + \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dB_s \\
&= u(\hat{Y}_0) - T \left(\frac{1}{T} \int_0^T f(X_s^\sigma) ds - \int f d\nu^\sigma \right) + \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dB_s
\end{aligned}$$

Thus,

$$\sqrt{T} \left(\frac{1}{T} \int_0^T f(X_s^\sigma) ds - \int f d\nu^\sigma \right) = \frac{1}{\sqrt{T}} \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dB_s - \frac{1}{\sqrt{T}} \left(u(\hat{Y}_T) - u(\hat{Y}_0) \right).$$

We aim to apply the Central Limit Theorem for stochastic integrals in order to prove the convergence in law. The bracket of the stochastic integral appearing in the expression is

$$\frac{1}{T} \int_0^T \left(\rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) \right)^2 ds.$$

We are now going to express this integral in terms of the process X^σ and apply the result of proposition 2.1.1 to check its almost sure convergence :

$$\begin{aligned}
\sigma_f^2 &:= \text{a.s.} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\rho(\hat{Y}_s) \cdot \frac{du}{dy}(\hat{Y}_s) \right)^2 ds \\
&= \text{a.s.} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\rho \circ \varphi_{p,q}(X_s^\sigma) \cdot \frac{du}{dy} \circ \varphi_{p,q}(X_s^\sigma) \right)^2 ds \\
&= \int_{\mathbb{R}} \left(\rho \circ \varphi_{p,q}(z) \cdot \frac{du}{dy} \circ \varphi_{p,q}(z) \right)^2 d\nu^\sigma(z) \\
&= 4 \int_{\mathbb{R}} \sigma(z)^2 \left(\int_z^b \frac{f(y) - \int f d\nu^\sigma}{\sigma(y)^2} dy \right)^2 d\nu^\sigma(z) < +\infty.
\end{aligned}$$

The assumptions of the Central Limit Theorem for Brownian martingales are verified (see Basawa and Rao [8]) and therefore

$$\frac{1}{\sqrt{T}} \int_0^T \rho(\hat{Y}_s) \frac{du}{dy}(\hat{Y}_s) dB_s \xrightarrow[T \rightarrow +\infty]{Law} \mathcal{N}(0, \sigma_f^2).$$

Moreover, as the function u is continuous over the compact set $[\varphi_{p,q}(a), \varphi_{p,q}(b)]$, it is bounded and so

$$\frac{1}{\sqrt{T}} \left(u(\hat{Y}_T) - u(\hat{Y}_0) \right) \xrightarrow[T \rightarrow +\infty]{\mathbb{P}} 0.$$

Finally, we apply Slutsky's lemma to deduce the convergence in law

$$\sqrt{T} \left(\frac{1}{T} \int_0^T f(X_s^\sigma) ds - \int f d\nu^\sigma \right) \xrightarrow[T \rightarrow +\infty]{Law} \mathcal{N}(0, \sigma_f^2).$$

□

We have now all the elements to prove the main theorem of this paper.

PROOF (of theorem 2.1.1). Remember the definition of the estimator $\hat{\alpha}_T$:

$$\hat{\alpha}_T := \frac{M_T \int_a^\gamma \frac{dx}{\sigma^2(x)}}{M_T \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) + (b - \gamma)},$$

with

$$M_T := \frac{1}{T} \int_0^T \sigma^2(X_s^\sigma) \mathbb{1}_{(X_s^\sigma \geq \gamma)} ds, \quad T > 0.$$

From proposition 2.1.1, we know that the process M converges almost surely,

$$M_T \xrightarrow[T \rightarrow +\infty]{a.s.} \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma.$$

This limit can be restated as a function of α :

$$\begin{aligned} \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma &= \frac{\int_{\mathbb{R}} \sigma^2(x) \mathbb{1}_{(x \geq \gamma)} \frac{d\nu^1(x)}{\sigma^2(x)}}{\int_{\mathbb{R}} \frac{d\nu^1(x)}{\sigma^2(x)}} \\ &= \frac{\alpha(b - \gamma)}{\alpha \int_\gamma^b \frac{dx}{\sigma^2(x)} + (1 - \alpha) \int_a^\gamma \frac{dx}{\sigma^2(x)}}. \end{aligned}$$

From this equation, we check the strong consistency of the estimator $\hat{\alpha}_T$. To prove the asymptotic normality, we have to rewrite $\hat{\alpha}_T$ in a slightly different way :

$$\begin{aligned} \sqrt{T}(\hat{\alpha}_T - \alpha) &= \sqrt{T} \frac{M_T \int_a^\gamma \frac{dx}{\sigma^2(x)} - \alpha M_T \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) - \alpha(b - \gamma)}{M_T \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) + (b - \gamma)} \\ &= \frac{\sqrt{T} \left((1 - \alpha) \int_a^\gamma \frac{dx}{\sigma^2(x)} + \alpha \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) \left(M_T - \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma \right)}{M_T \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) + (b - \gamma)} \end{aligned}$$

Using the results of the propositions 2.1.1 and 2.1.2, we know that, on the one hand,

$$M_T \xrightarrow[T \rightarrow +\infty]{a.s.} \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma,$$

and, on the other hand,

$$\sqrt{T} \left(M_T - \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma \right) \xrightarrow[T \rightarrow +\infty]{\text{Law}} \mathcal{N}(0, \tau^2),$$

with

$$\tau^2 := 4 \int_{\mathbb{R}} \sigma(z)^2 \left(\int_z^b \frac{\sigma^2(y) \mathbb{1}_{(y \geq \gamma)} - \int_\gamma^{+\infty} \sigma^2 d\nu^\sigma}{\sigma(y)^2} dy \right)^2 d\nu^\sigma(y).$$

Therefore, applying Slutsky's lemma again, we can conclude that our estimator is asymptotically normal :

$$\sqrt{T} (\hat{\alpha}_T - \alpha) \xrightarrow[T \rightarrow +\infty]{\text{Law}} \mathcal{N} (0, \kappa^2)$$

with variance

$$\kappa^2 := \left(\frac{\left((1 - \alpha) \int_a^\gamma \frac{dx}{\sigma^2(x)} + \alpha \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) \tau}{\int_\gamma^{+\infty} \sigma^2 d\nu^\sigma \left(\int_a^\gamma \frac{dx}{\sigma^2(x)} - \int_\gamma^b \frac{dx}{\sigma^2(x)} \right) + (b - \gamma)} \right)^2.$$

This ends the proof of our main theorem. \square

2.3. Process with drift. We now turn to the general case of skew processes with drift. Let X , be the unique weak solution to the stochastic differential equation :

$$(100) \quad X_t = x + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s + (2\alpha - 1)L_t^\gamma(X) - K_t^{[a,b]}(X), \quad x \in [a, b], \quad \alpha \in]0, 1[.$$

The function b is supposed to be Lipschitz, and, once more, the functions b and σ are supposed to be known.

A scale function of the symmetric process ($\alpha = 0$) is

$$s(y) := \int_a^y \exp \left(-2 \int_a^z \frac{b(u)}{\sigma^2(u)} du \right) dy, \quad a \leq y \leq b.$$

As the function s is twice differentiable, the Itô-Tanaka formula applied to this function and the process X leads to the following representation for the process $X^s := s(X)$.

$$X_t^s = s(x) + \int_0^t (\sigma s') \circ s^{-1}(X_u^s) dB_u + (2\alpha - 1)L_t^{s(\gamma)}(X^s) - \int_0^t s' \circ s^{-1}(X_u^s) dK_u^{[a,b]}(X).$$

Now, let us define the function $\sigma^s := (\sigma s') \circ s^{-1}$ and the process

$$K_t^{[s(a), s(b)]}(X^s) := \int_0^t s' \circ s^{-1}(X_u^s) dK_u^{[a,b]}(X).$$

One can easily check that the process X^s satisfies the following SDE in the weak sense :

$$X_t^s = s(x) + \int_0^t \sigma^s(X_u^s) dB_u + (2\alpha - 1)L_t^{s(\gamma)}(X^s) - K_t^{[s(a), s(b)]}(X^s),$$

where the process $K^{[s(a), s(b)]}(X^s)$ induces a reflection over the interval $[s(a), s(b)]$. As the process X is reflected over the interval $[s(a), s(b)]$ this equation is precisely the SDE (83) studied in the first part of this paper. Following exactly the same demonstration, it is then a trivial matter to prove that

THEOREM 2.3.1. *The estimator of the parameter α defined by*

$$\hat{\alpha}_T := \frac{M_T \int_a^\gamma \frac{dx}{(\sigma^2 s')(x)}}{M_T \left(\int_a^\gamma \frac{dx}{(\sigma^2 s')(x)} - \int_\gamma^b \frac{dx}{(\sigma^2 s')(x)} \right) + s(b)}, \quad T > 0,$$

where

$$M_T := \frac{1}{T} \int_0^T (\sigma s')^2(X_s) \mathbb{1}_{(X_s \geq \gamma)} ds, \quad T > 0,$$

is strongly consistent and asymptotically normal.

In particular, the strong consistency is a consequence of the following ergodic theorem.

THEOREM 2.3.2. *Let us define the new measure*

$$\mu(dx) \propto (\alpha p(x) \mathbb{1}_{[\gamma, b]}(x) + (1 - \alpha) p(x) \mathbb{1}_{[a, \gamma]}(x)) dx$$

with

$$p(x) \propto \frac{\exp\left(2 \int_a^x \frac{b(z)}{\sigma^2(z)} dz\right)}{\sigma(x)^2}, \quad \forall x \in [a, b].$$

Then, for any function $f \in L^1(\mu)$,

$$\frac{1}{T} \int_0^T f(X_s) ds \xrightarrow[T \rightarrow +\infty]{a.s.} \int f d\mu.$$

The proofs are similar to the case $b \equiv 0$. They rely on changes of variable using the bijection s (see for example Rogers and Williams [78] for this proof in the case $\alpha = \frac{1}{2}$).

3. Estimation of the parameters α and γ when they are both unknown

We are now going to explain how one can construct an efficient estimator of the parameter γ as a minimum contrast estimator. First, we need a few results about time changed diffusion processes.

3.1. Random time change and reflected diffusions. We shall make use of the following lemma, extending a well known result concerning the brownian motion (see Karatzas and Shreve [47]).

LEMMA 3.1.1. If the process X is solution to the following SDE

$$(101) \quad X_t = x + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s + (2\alpha - 1)L_t^\gamma(X) - K_t^{[a, b]}(X),$$

for $\theta \in]a, b[$, we define the random time change

$$\tilde{\tau}_t^\theta := \inf \left\{ s \geq 0 : \int_0^s \mathbb{1}_{(X_s \geq \theta)} ds > t \right\}.$$

Then, $\check{X}_t^\theta := X_{\tilde{\tau}_t^\theta}$ is solution to the SDE

$$\check{X}_t^\theta = \check{X}_0^\theta + \int_0^t b(\check{X}_s^\theta) ds + \int_0^t \sigma(\check{X}_s^\theta) d\check{W}_s + (2\alpha - 1)L_t^\gamma(\check{X}^\theta) - K_t^{[\theta, b]}(\check{X}^\theta)$$

if $\gamma \geq \theta$ and

$$\check{X}_t^\theta = \check{X}_0^\theta + \int_0^t b(\check{X}_s^\theta) ds + \int_0^t \sigma(\check{X}_s^\theta) d\check{W}_s - K_t^{[\theta, b]}(\check{X}^\theta)$$

if $\gamma \leq \theta$, where \check{W} is a new standard brownian motion and $\check{X}_0^\theta := \inf_{s \geq 0} \{X_s \geq \theta\}$.

PROOF. From the definition of the stopping time $\tilde{\tau}_t^\theta$, it is clear that, on the one hand,

$$\int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} ds = \int_0^t \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} d\tilde{\tau}_s^\theta = t$$

and, on the other hand,

$$\int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dB_s = \int_0^t \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} dB_{\tilde{\tau}_s^\theta}$$

is a standard brownian motion \check{W}_t from the Dubbins-Schwarz theorem. Moreover, by construction, $\check{X}_t^\theta \geq \theta$, almost surely, for all $t \geq 0$. Thus,

$$\begin{aligned} \check{X}_t^\theta - \theta &= |\check{X}_t^\theta - \theta| \text{ as } \check{X}_t^\theta \geq \theta \\ &= |\check{X}_0^\theta - \theta| + \int_0^t \text{sgn}(\check{X}_s^\theta - \theta) d\check{X}_s^\theta + L_t^\theta(\check{X}^\theta) \text{ from the Tanaka formula} \\ &= |\check{X}_0^\theta - \theta| + \int_0^t \mathbb{1}_{(\check{X}_s^\theta \geq \theta)} d\check{X}_s^\theta + L_t^\theta(\check{X}^\theta) \text{ because } \check{X}_t^\theta \geq \theta \\ &= \check{X}_0^\theta - \theta + \int_0^t \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} dX_{\tilde{\tau}_s^\theta} + L_t^\theta(\check{X}^\theta) \text{ because } \check{X}_0^\theta \geq \theta \end{aligned}$$

Thus, we get the first representation

$$\check{X}_t^\theta = \check{X}_0^\theta + \int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dX_s + L_t^\theta(\check{X}^\theta)$$

We now make use of equation (101) :

$$\begin{aligned} \int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dX_s &= \int_0^{\tilde{\tau}_t^\theta} b(X_s) \mathbb{1}_{(X_s \geq \theta)} ds + \int_0^{\tilde{\tau}_t^\theta} \sigma(X_s) \mathbb{1}_{(X_s \geq \theta)} dB_s \\ &\quad + (2\alpha - 1) \int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dL_s^\gamma(X) - \int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dK_s^{[a, b]}(X). \end{aligned}$$

Considering the first two integrals and our preliminary remarks, we have, on the one hand

$$\begin{aligned} \int_0^{\tilde{\tau}_t^\theta} b(X_s) \mathbb{1}_{(X_s \geq \theta)} ds &= \int_0^t b(X_{\tilde{\tau}_s^\theta}) \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} d\tilde{\tau}_s^\theta \\ &= \int_0^t b(\tilde{X}_s^\theta) ds, \end{aligned}$$

and, on the other hand,

$$\begin{aligned} \int_0^{\tilde{\tau}_t^\theta} \sigma(X_s) \mathbb{1}_{(X_s \geq \theta)} dB_s &= \int_0^t \sigma(X_{\tilde{\tau}_s^\theta}) \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} dB_{\tilde{\tau}_s^\theta} \\ &= \int_0^t \sigma(\tilde{X}_s^\theta) d\tilde{W}_s. \end{aligned}$$

Consider now the term

$$\begin{aligned} \int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dL_s^\gamma(X) &= \int_0^t \mathbb{1}_{(X_{\tilde{\tau}_s^\theta} \geq \theta)} dL_{\tilde{\tau}_s^\theta}^\gamma(X) \\ &= \int_0^t \mathbb{1}_{(\tilde{X}_s^\theta \geq \theta)} dL_s^\gamma(\tilde{X}^\theta). \end{aligned}$$

It is the clear, from the property of the local time that

$$\begin{aligned} - \int_0^t \mathbb{1}_{(\tilde{X}_s^\theta \geq \theta)} dL_s^\gamma(\tilde{X}^\theta) &= 0 \text{ if } \gamma \leq \theta, \\ - \int_0^t \mathbb{1}_{(\tilde{X}_s^\theta \geq \theta)} dL_s^\gamma(\tilde{X}^\theta) &= L_t^\gamma(\tilde{X}^\theta) \text{ if } \gamma \geq \theta. \end{aligned}$$

For the same reasons, we also have

$$\int_0^{\tilde{\tau}_t^\theta} \mathbb{1}_{(X_s \geq \theta)} dK_s^{[a, b]}(X) = K_t^{[\theta, b]}(\tilde{X}^\theta).$$

And the proof is complete. \square

The graphics 2, 3 and 4 illustrate the former construction for a brownian motion. In such a case, one can find an alternative demonstration, based on the same ideas, in Itô and McKean [45] or Stroock and Varadhan [81].

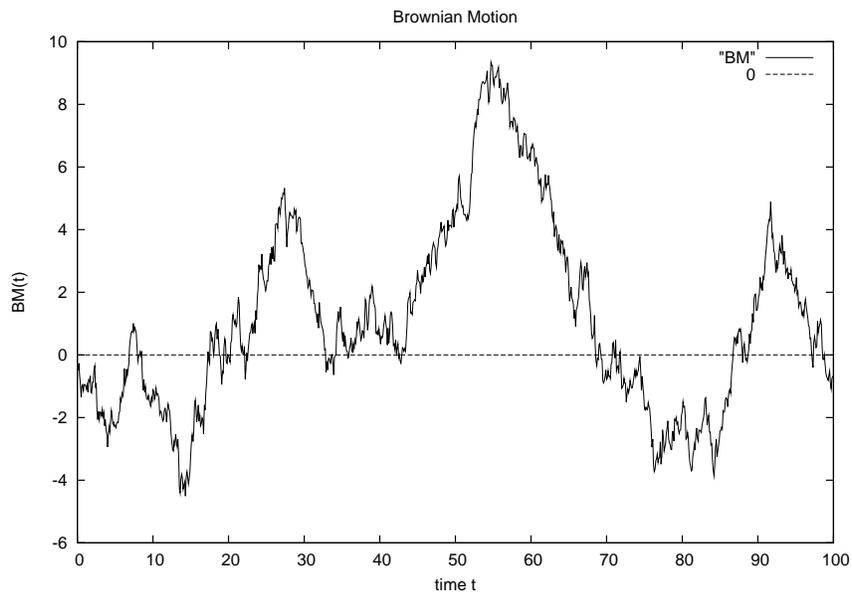


FIG. 2. Brownian motion X_t

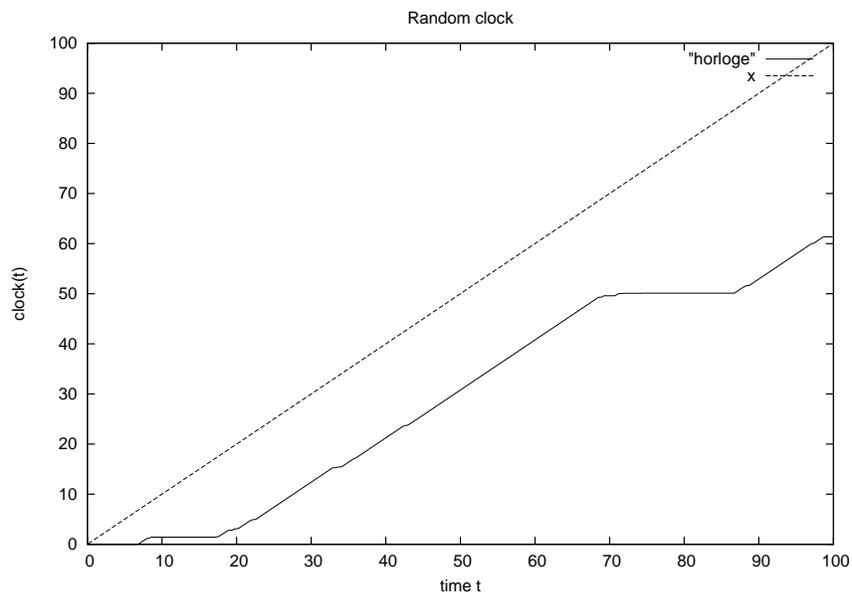


FIG. 3. Random time change $t \mapsto \int_0^t \mathbb{1}_{(X_s \geq 0)} ds$

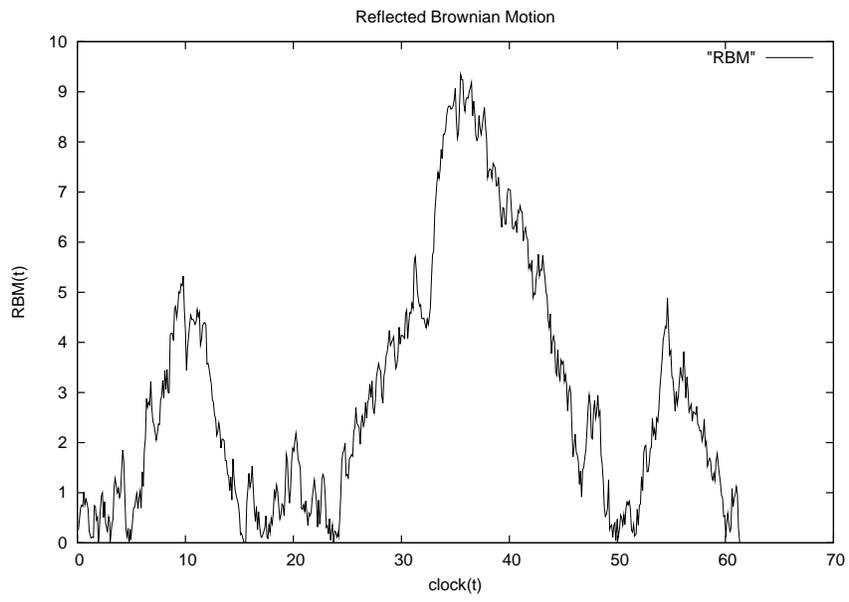


FIG. 4. Reflected brownian motion \check{X}_t^0

REMARK. Of course, defining the time change

$$\hat{\tau}_t^\theta := \inf \left\{ s \geq 0 : \int_0^s \mathbb{1}_{(X_s \leq \theta)} ds > t \right\},$$

it is easily seen that the process $\hat{X}_t^\theta := X_{\hat{\tau}_t^\theta}$ will be a solution to the following SDE

$$\hat{X}_t^\theta = \hat{X}_0^\theta + \int_0^t b(\hat{X}_s^\theta) ds + \int_0^t \sigma(\hat{X}_s^\theta) d\hat{W}_s + (2\alpha - 1)L_t^\gamma(\hat{X}^\theta) - K_t^{[a, \theta]}(\hat{X}^\theta)$$

if $\gamma \leq \theta$ and

$$\hat{X}_t^\theta = \hat{X}_0^\theta + \int_0^t b(\hat{X}_s^\theta) ds + \int_0^t \sigma(\hat{X}_s^\theta) d\hat{W}_s - K_t^{[a, \theta]}(\hat{X}^\theta)$$

if $\gamma \geq \theta$, where \hat{W} is a new standard brownian motion and $\hat{X}_0^\theta := \inf_{s \geq 0} \{X_s \leq \theta\}$.

For convenience, we now define the right inverses of the random times $\check{\tau}_t^\theta$ and $\hat{\tau}_t^\theta$:

$$\check{f}(t, \theta) := \int_0^t \mathbb{1}_{(X_s \geq \theta)} ds$$

and

$$\hat{f}(t, \theta) := \int_0^t \mathbb{1}_{(X_s \leq \theta)} ds.$$

We now prove a lemma concerning the regularity of the random time changes.

LEMMA 3.1.2. For all $t \geq 0$, the random functions $\theta \mapsto \check{\tau}_t^\theta$ and $\theta \mapsto \hat{\tau}_t^\theta$ are differentiable and

$$\begin{aligned} \frac{\partial}{\partial \theta} \check{\tau}_t^\theta &= \frac{L_t^\theta(\check{X}^\theta)}{\sigma^2(\check{X}_t^\theta)}, \\ \frac{\partial}{\partial \theta} \hat{\tau}_t^\theta &= -\frac{L_t^\theta(\hat{X}^\theta)}{\sigma^2(\hat{X}_t^\theta)}. \end{aligned}$$

PROOF. We prove the lemma for the map $\theta \mapsto \check{\tau}_t^\theta$ as the demonstration is similar in the other case. For $\theta \in]a, b[$ and $t > 0$, we have

$$\begin{aligned} \check{f}(t, \theta) &= \int_0^t \mathbb{1}_{(X_s \geq \theta)} ds \\ &= \int_\theta^{+\infty} \frac{L_s^y(X)}{\sigma^2(X_s)} dy \text{ from the occupation time formula.} \end{aligned}$$

From the first expression, we see that $t \mapsto \check{f}(t, \theta)$ is differentiable for all $\theta \in]a, b[$ with

$$\frac{\partial}{\partial t} \check{f}(t, \theta) = \mathbb{1}_{(X_t \geq \theta)},$$

and from the second, that $\theta \mapsto \check{f}(t, \theta)$ is differentiable for all $t > 0$ with

$$\frac{\partial}{\partial \theta} \check{f}(t, \theta) = -\frac{L_t^\theta(X)}{\sigma^2(X_t)}.$$

Moreover, by definition, the map $t \mapsto \check{\tau}_t^\theta$ is the right inverse of $t \mapsto \check{f}(t, \theta)$. Thus, we have

$$\check{f}(\check{\tau}_t^\theta, \theta) = t, \forall t > 0.$$

Then, from the implicit functions theorem, $\theta \mapsto \check{\tau}_t^\theta$ is differentiable and

$$\frac{\partial}{\partial \theta} \check{\tau}_t^\theta = -\frac{\frac{\partial}{\partial \theta} \check{f}(\check{\tau}_t^\theta, \theta)}{\frac{\partial}{\partial t} \check{f}(\check{\tau}_t^\theta, \theta)}.$$

But, we also have, on the one hand

$$\begin{aligned} \frac{\partial}{\partial t} \check{f}(\check{\tau}_t^\theta, \theta) &= \mathbb{1}_{(X_{\check{\tau}_t^\theta} \geq \theta)} \\ &= \mathbb{1}_{(\check{X}_t^\theta \geq \theta)} \\ &= 1 \end{aligned}$$

and, on the other hand,

$$\begin{aligned} \frac{\partial}{\partial \theta} \check{f}(\check{\tau}_t^\theta, \theta) &= -\frac{L_{\check{\tau}_t^\theta}^\theta(X)}{\sigma^2(X_{\check{\tau}_t^\theta})} \\ &= -\frac{L_t^\theta(\check{X}^\theta)}{\sigma^2(\check{X}_t^\theta)}. \end{aligned}$$

This proves the lemma. □

3.2. Estimator of γ .

3.2.1. *Main idea.* The idea to estimate γ is very simple. We are going to split the interval $[a, b]$ at a given level θ and stick each part of the sample path of X using the time changes described above. Then, a fictive asymmetry coefficient will be estimated over each one of the sub-intervals $[a, \theta]$ and $[\theta, b]$. If the splitting occurs precisely at the level $\theta = \gamma$, then no asymmetry should be detected and the estimators of the fictive asymmetry coefficients should be equal to $\frac{1}{2}$. Let us now make precise this intuitive idea.

3.2.2. *The estimator of γ and its consistency.* Keeping the same notations, we now define, for $a < \theta < b$, the empirical means

$$\check{M}_T^\theta := \frac{1}{\check{f}(T, \theta)} \int_0^{\check{f}(T, \theta)} (\sigma s'_\theta)^2 (\check{X}_s^\theta) \mathbb{1}_{(\check{X}_s^\theta \geq \frac{\theta+b}{2})} ds$$

and

$$\hat{M}_T^\theta := \frac{1}{\hat{f}(T, \theta)} \int_0^{\hat{f}(T, \theta)} (\sigma s'_a)^2 (\hat{X}_s^\theta) \mathbb{1}_{(\hat{X}_s^\theta \geq \frac{\theta+a}{2})} ds,$$

where $s(\cdot)$ is the function

$$s_z(y) := \int_z^y \exp\left(-2 \int_a^z \frac{b(u)}{\sigma^2(u)} du\right) dy, \quad a \leq z \leq y \leq b.$$

Unsurprisingly, we also define the corresponding estimators

$$\check{\alpha}_T^\theta := \frac{\check{M}_T^\theta \int_\theta^{\frac{b+\theta}{2}} \frac{dx}{(\sigma^2 s'_\theta)(x)}}{\check{M}_T^\theta \left(\int_\theta^{\frac{b+\theta}{2}} \frac{dx}{(\sigma^2 s'_\theta)(x)} - \int_{\frac{b+\theta}{2}}^b \frac{dx}{(\sigma^2 s'_\theta)(x)} \right) + s_\theta(b)}$$

and

$$\hat{\alpha}_T^\theta := \frac{\hat{M}_T^\theta \int_a^{\frac{a+\theta}{2}} \frac{dx}{(\sigma^2 s'_a)(x)}}{\hat{M}_T^\theta \left(\int_a^{\frac{a+\theta}{2}} \frac{dx}{(\sigma^2 s'_a)(x)} - \int_{\frac{a+\theta}{2}}^\theta \frac{dx}{(\sigma^2 s'_a)(x)} \right) + s_a(\theta)}.$$

Finally, the estimator of the parameter γ is defined by

$$\hat{\gamma}_T \in \operatorname{argmin}_{a < \theta < b} U_T(\theta)$$

with

$$U_T(\theta) := |\check{\alpha}_T^\theta - \hat{\alpha}_T^\theta|.$$

The main result of this section is the following theorem.

THEOREM 3.2.1. *For $\alpha \neq \frac{1}{2}$, the random variable $\hat{\gamma}_T$ is a strongly consistent estimator of the parameter γ .*

To prove this theorem, we show that U_T is a contrast process converging towards the contrast function U with a unique minimum in γ .

PROOF. Remember that the invariant density of the process X is

$$p(y) \propto \alpha \frac{\mathbb{1}_{[\gamma, b]}(y)}{(\sigma^2 s'_a)(y)} + (1 - \alpha) \frac{\mathbb{1}_{[a, \gamma]}(y)}{(\sigma^2 s'_a)(y)}$$

so that

$$\frac{L_T^y(X)}{T} \xrightarrow[T \rightarrow +\infty]{\text{a.s.}} \sigma^2(y)p(y), \forall y \in [a, b] \text{ (see Bosq and Davydov [16] or Van Zanten[88] for instance).}$$

Using the time changes defined previously and the occupation time formula, we have the alternative expressions

$$\check{M}_T^\theta = \frac{\int_{\frac{\theta+b}{2}}^b s'_\theta(y)^2 L_T^y(X) dy}{\int_\theta^b \frac{L_T^y(X)}{\sigma^2(y)} dy},$$

$$\hat{M}_T^\theta = \frac{\int_{\frac{\theta+a}{2}}^\theta s'_a(y)^2 L_T^y(X) dy}{\int_a^\theta \frac{L_T^y(X)}{\sigma^2(y)} dy},$$

and the convergence results

$$\check{M}_T^\theta \xrightarrow[T \rightarrow +\infty]{\text{a.s.}} \check{M}^\theta := \frac{\int_{\frac{\theta+b}{2}}^b (\sigma s'_\theta)^2(y) p(y) dy}{\int_\theta^b p(y) dy},$$

$$\hat{M}_T^\theta \xrightarrow[T \rightarrow +\infty]{\text{a.s.}} \hat{M}^\theta := \frac{\int_{\frac{\theta+a}{2}}^\theta (\sigma s'_a)^2(y) p(y) dy}{\int_a^\theta p(y) dy}.$$

From the previous equations, it is now easy to check that U_T can be written

$$U_T(\theta) = |\check{\alpha}_T^\theta - \hat{\alpha}_T^\theta|$$

with

$$\check{\alpha}_T^\theta := \frac{\int_{\frac{\theta+b}{2}}^b s'_\theta(y)^2 L_T^y(X) dy \int_\theta^{\frac{b+\theta}{2}} \frac{dy}{(\sigma^2 s'_\theta(y))}}{\int_{\frac{\theta+b}{2}}^b s'_\theta(y)^2 L_T^y(X) dy \left(\int_\theta^{\frac{b+\theta}{2}} \frac{dy}{(\sigma^2 s'_\theta(y))} - \int_{\frac{b+\theta}{2}}^b \frac{dy}{(\sigma^2 s'_\theta(y))} \right) + s_\theta(b) \int_\theta^b \frac{L_T^y(X)}{\sigma^2(y)} dy},$$

and

$$\hat{\alpha}_T^\theta := \frac{\int_{\frac{\theta+a}{2}}^\theta s'_a(y)^2 L_T^y(X) dy \int_a^{\frac{a+\theta}{2}} \frac{dy}{(\sigma^2 s'_a(y))}}{\int_{\frac{\theta+a}{2}}^\theta s'_a(y)^2 L_T^y(X) dy \left(\int_a^{\frac{a+\theta}{2}} \frac{dy}{(\sigma^2 s'_a(y))} - \int_{\frac{a+\theta}{2}}^\theta \frac{dy}{(\sigma^2 s'_a(y))} \right) + s_a(\theta) \int_a^\theta \frac{L_T^y(X)}{\sigma^2(y)} dy}.$$

Moreover, for all $\theta \in]a, b[$, $U_T(\theta)$ converges almost surely pointwise towards the deterministic function

$$U(\theta) := |\check{\alpha}^\theta - \hat{\alpha}^\theta|$$

with

$$\check{\alpha}^\theta := \frac{\int_{\frac{\theta+b}{2}}^b (\sigma s'_\theta)^2(y) p(y) dy \int_\theta^{\frac{b+\theta}{2}} \frac{dy}{(\sigma^2 s'_\theta(y))}}{\int_{\frac{\theta+b}{2}}^b (\sigma s'_\theta)^2(y) p(y) dy \left(\int_\theta^{\frac{b+\theta}{2}} \frac{dy}{(\sigma^2 s'_\theta(y))} - \int_{\frac{b+\theta}{2}}^b \frac{dy}{(\sigma^2 s'_\theta(y))} \right) + s_\theta(b) \int_\theta^b p(y) dy},$$

and

$$\hat{\alpha}^\theta := \frac{\int_{\frac{\theta+a}{2}}^\theta (\sigma s'_a)^2(y) p(y) dy \int_a^{\frac{a+\theta}{2}} \frac{dy}{(\sigma^2 s'_a(y))}}{\int_{\frac{\theta+a}{2}}^\theta (\sigma s'_a)^2(y) p(y) dy \left(\int_a^{\frac{a+\theta}{2}} \frac{dy}{(\sigma^2 s'_a(y))} - \int_{\frac{a+\theta}{2}}^\theta \frac{dy}{(\sigma^2 s'_a(y))} \right) + s_a(\theta) \int_a^\theta p(y) dy}.$$

Now, the study of the function $\theta \mapsto U(\theta)$ easily shows that it is continuous and admits a unique minimum at $\theta = \gamma$, for all $\alpha \neq \frac{1}{2}$. The figures 5, 6 and 7 represent this function for different values of the parameters in the case of the skew brownian motion. For this process, we have $\sigma \equiv 1$ and $s_z(y) = y - z$. In this setting, we get the more tractable expressions

$$\check{\alpha}_T^\theta = \frac{1}{2} \frac{\int_{\frac{b+\theta}{2}}^b L_T^y dy}{\int_\theta^b L_T^y dy} \xrightarrow[T \rightarrow +\infty]{\text{a.s.}} \frac{1}{2} \frac{\alpha (b - \gamma \vee \frac{b+\theta}{2}) + (1 - \alpha) (\gamma - \frac{b+\theta}{2})_+}{\alpha (b - \gamma \vee \theta) + (1 - \alpha) (\gamma - \theta)_+}$$

and

$$\begin{aligned} \hat{\alpha}_T^\theta &= \frac{1}{2} \frac{\int_{\frac{a+\theta}{2}}^\theta L_T^y dy}{\int_a^\theta L_T^y dy} \xrightarrow[T \rightarrow +\infty]{\text{a.s.}} \frac{1}{2} \frac{\frac{\theta-a}{2} \left(\alpha \mathbb{1}_{(\theta \leq \gamma)} + (1 - \alpha) \mathbb{1}_{(\gamma \leq \frac{\theta+a}{2})} \right)}{\alpha (\theta - \gamma)_+ + (1 - \alpha) (\theta \wedge \gamma - a)} \\ &\quad + \frac{(\alpha (\theta - \gamma) + (1 - \alpha) (\gamma - \frac{\theta+a}{2})) \mathbb{1}_{(\frac{\theta+a}{2} \leq \gamma \leq \theta)}}{\alpha (\theta - \gamma)_+ + (1 - \alpha) (\theta \wedge \gamma - a)} \end{aligned}$$

where L_T^y stands for $L_T^y(X)$.

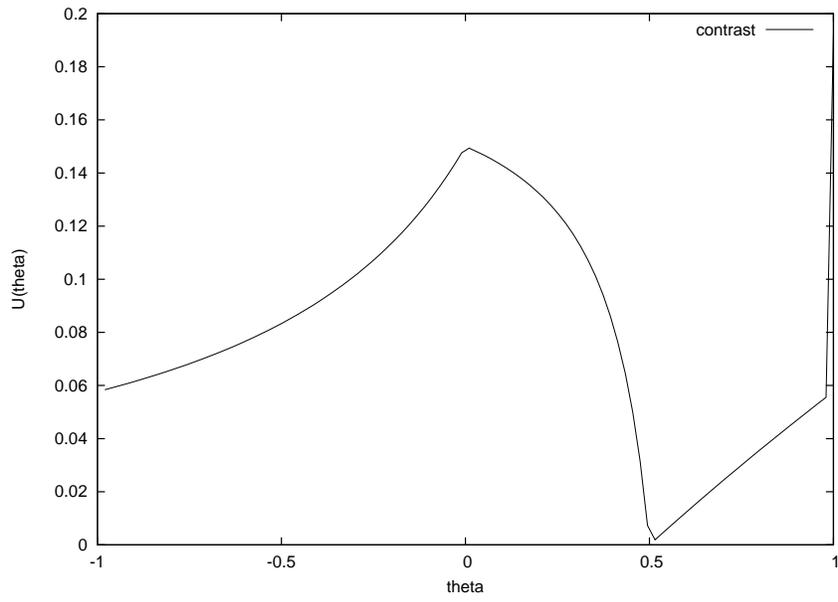


FIG. 5. $\theta \mapsto U(\theta)$ for $a = -1.0$, $b = 1.0$, $\alpha = 0.2$ and $\gamma = 0.5$

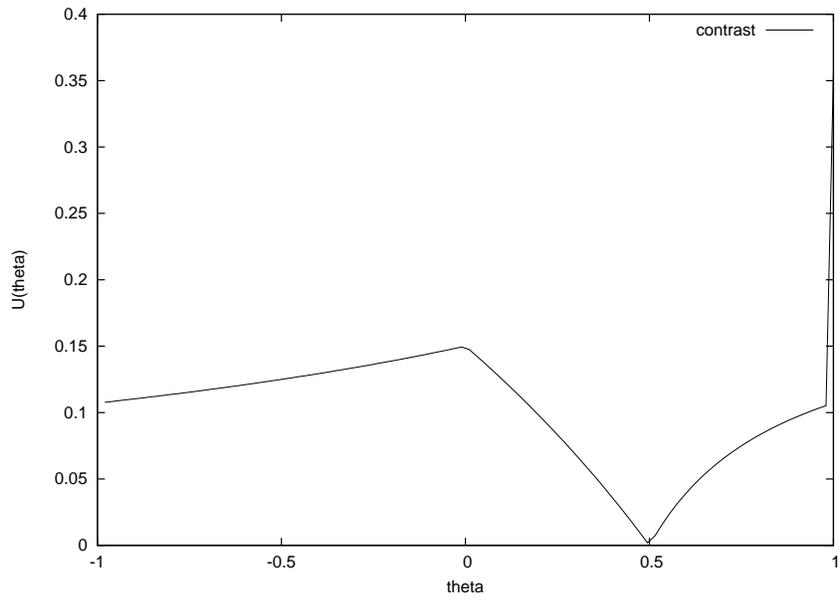


FIG. 6. $\theta \mapsto U(\theta)$ for $a = -1.0$, $b = 1.0$, $\alpha = 0.8$ and $\gamma = 0.5$

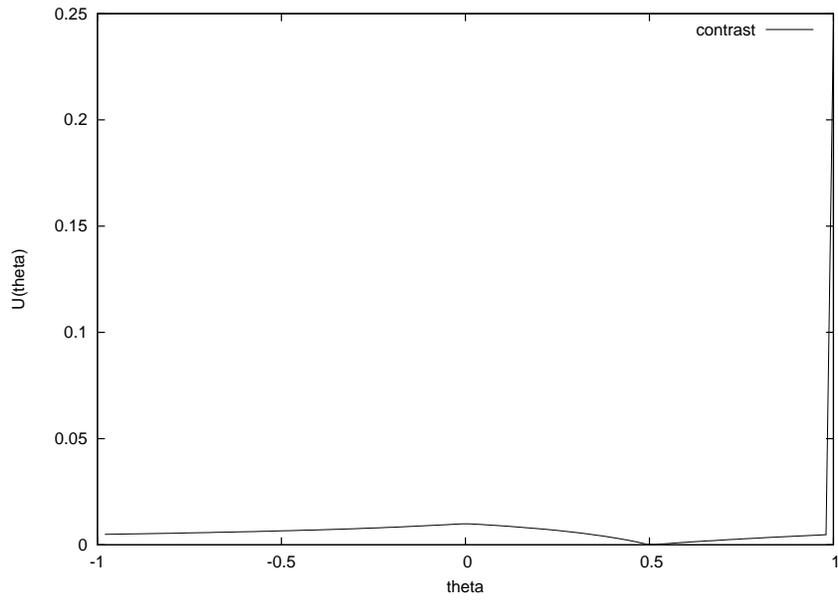


FIG. 7. $\theta \mapsto U(\theta)$ for $a = -1.0$, $b = 1.0$, $\alpha = 0.48$ and $\gamma = 0.5$

Therefore, we have checked that $U_T(\theta)$ is a contrast process. To our knowledge, there is no definitive result about the convergence of argmin estimators. In the general case, the good notion seems to be the epi-convergence of sequence of functions as studied by C. Hess [43]. But our situation is very simple. In fact, our contrast process defines a sequence of continuous functions defined on a compact interval and converging almost surely pointwise to a continuous function with a unique minimum. Therefore, we have uniform convergence of our contrast process and thus, convergence of the argmin. This concludes the proof. \square

4. Numerical results

This last section illustrates the accuracy of the estimators on the exemple of the skew brownian motion starting from 0.

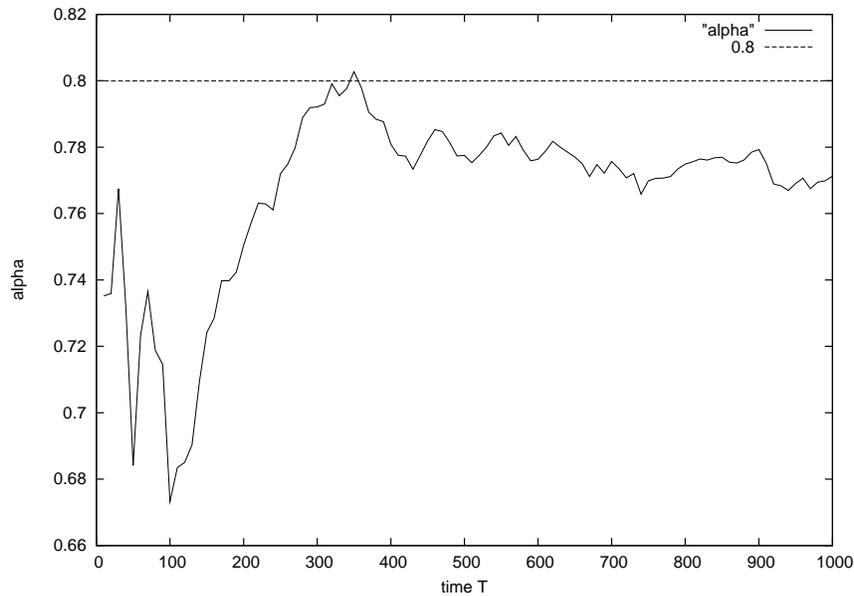


FIG. 8. Estimation of α when $a = -1.0$, $b = 1.0$, $\alpha = 0.8$, $\gamma = 0.0$

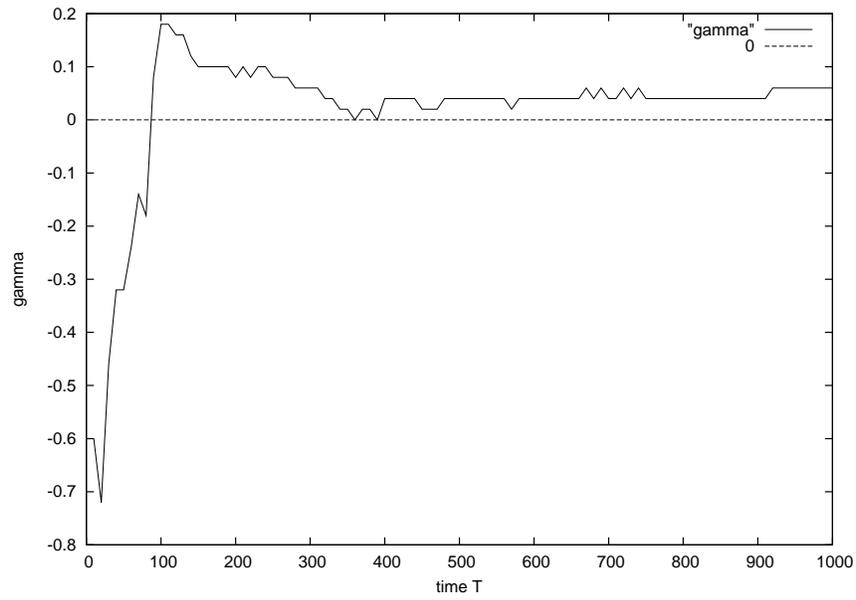


FIG. 9. Estimation of γ when $a = -1.0$, $b = 1.0$, $\alpha = 0.8$, $\gamma = 0.0$

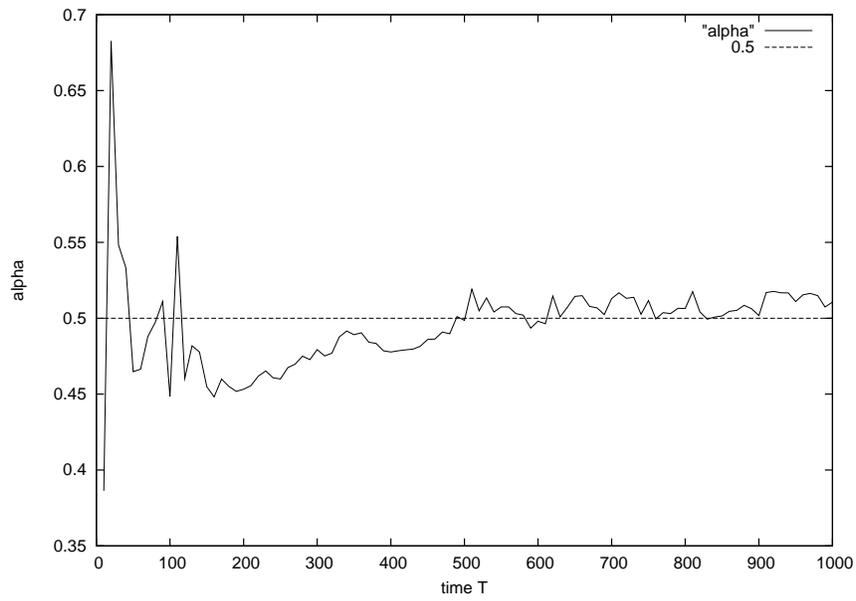


FIG. 10. Estimation of α when $a = -1.0$, $b = 1.0$, $\alpha = 0.5$, $\gamma = 0.0$

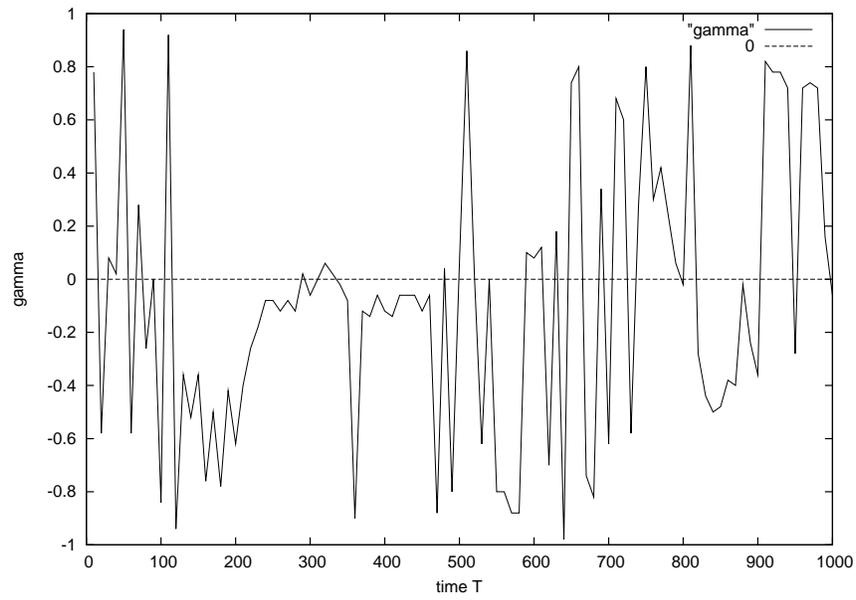


FIG. 11. Estimation of γ when $a = -1.0$, $b = 1.0$, $\alpha = 0.5$, $\gamma = 0.0$

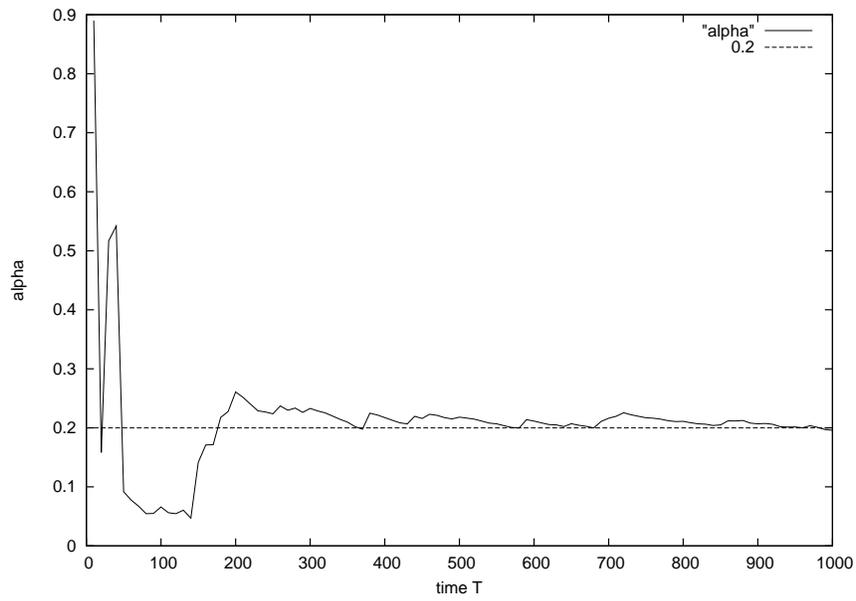


FIG. 12. Estimation of α when $a = -2.0$, $b = 2.0$, $\alpha = 0.2$, $\gamma = 1.0$

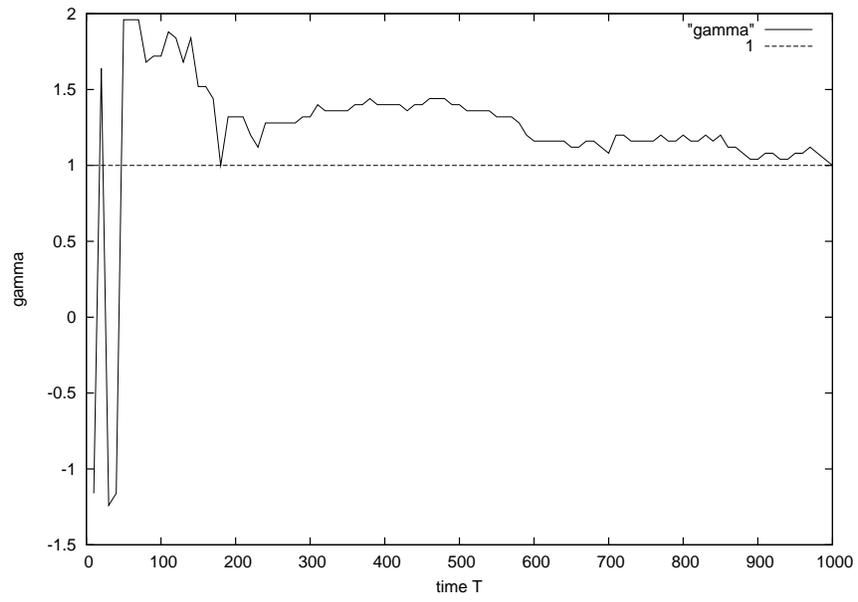


FIG. 13. Estimation of γ when $a = -2.0$, $b = 2.0$, $\alpha = 0.2$, $\gamma = 1.0$

5. Conclusion

5.1. Discussion upon our approach. It is worth pointing out that, as the measures of the processes $X(\alpha, \gamma)$ are singular for different values of α and γ , it is theoretically possible, in our continuous setting, to built estimators of the parameters with an exponential speed of convergence. Nevertheless, in practice, one has to deal with discrete data. Thus, such estimators, which are founded on the estimation of the local time at every point of the domain, may behave very badly. Meanwhile, our approach is robust to discretization. So, if the long run is not theoretically necessary in this estimation problem, we do think that in practice, such an asymptotic has to be considered.

5.2. Perspectives. This chapter is a first attempt to estimate the parameters of skewed processes. So far, we have provided a method to consistently estimate both the asymmetry coefficient and the locus of the asymmetry, and given a speed of convergence for these estimators, in a one dimensional setting. Much remains to do. Of course, the extension to the multidimensional setting is the great matter. But the lack of uniqueness of the notion of local time in higher dimension seems to forbid a general treatment of this question. Another way to explore is the study of kernel estimators to handle the functional case where the asymmetry is defined by a functional as $\int_0^t \alpha(X_s) dL_s^\gamma$ or $\alpha(L_t^\gamma)$ (see Barlow, Burdzy, Kaspi and Mandelbaum [7] for an introduction to such processes).

Troisième partie

Collaboration industrielle

Outre le travail théorique présenté dans les précédents chapitres, cette thèse fût également l'occasion de considérer des problèmes de recherche industrielle au cours d'une collaboration avec l'équipe PRIME de la société Gaz de France.

Ce travail a été réalisé en collaboration avec Axel Grorud (Université de Provence et CNRS), Antoine Lejay (INRIA), Love Lindholm (Ecole Polytechnique), Charlotte Poudroux (ESSI) et Denis Talay (INRIA). Quatre rapports techniques détaillent les différents résultats que nous allons résumer dans ce chapitre.

One of the chief duties of the mathematician in acting as an adviser to scientists is to discourage them from expecting too much from mathematics. (Norbert Wiener)

MODELISATION ET SIMULATION DES INDICES PETROLIERS

1. Introduction

1.1. Le problème posé par Gaz de France au Projet OMEGA. GDF est confronté à des problèmes de modélisation et de simulation de contrats d'achats à long terme de gaz. Les prix sont indexés sur divers indices économiques, en particulier des taux de change et des prix de produits pétroliers. Dans la littérature académique, il n'est pas actuellement proposé de modélisation convaincante de ces contrats qui permette de bien poser et de résoudre numériquement la question de leur évaluation. Le projet OMEGA s'est donc attaché à la résolution de ces questions.

1.2. Analyse économique sommaire. Notre approche est empirique et notre modèle est construit à partir de l'étude de données numériques fournies par Gaz de France et non sur la base de considérations économiques. Néanmoins, pour nous permettre de comprendre les mécanismes de formation des prix du pétrole et du gaz, nous nous sommes fondés sur un document technique de D. Babusiaux [4]. Celui-ci décrit les différents facteurs qui influencent la formation des prix sur le long, le moyen et le court terme.

Sur le long terme, la loi d'Hotelling, qui stipule que le prix d'une ressource épuisable doit croître à un taux égal au taux d'actualisation, reste encore un modèle de référence. Toutefois, des progrès techniques constants permettent aujourd'hui d'exploiter des ressources nouvelles et plus que la rareté des hydrocarbures, il semble que ce soient désormais les coûts techniques et sociaux (en particulier les nouvelles contraintes de respect de l'environnement) qui imposent la progression des prix à long terme. Malgré tout, notre étude a permis de mettre en évidence une certaine stationnarité des cours qui n'est pas sans rappeler le comportement des taux d'intérêt.

Sur le moyen terme, le cartel des pays producteurs de pétrole (avec à leur tête l'Arabie Saoudite) et le marché se partagent la détermination du prix du brut. Le premier peut généralement maintenir un prix donné dans une fourchette raisonnable. Si les prix devaient devenir trop élevés, les forces de rappel du marché joueraient pour le faire redescendre dans cette fourchette.

Sur le court terme, les marchés à terme assurent la fluidité et la transparence des marchés. Ils amplifient également les tendances observées chez les opérateurs pétroliers, notamment en ce qui concerne la gestion de leurs stocks.

Enfin, tous ces éléments, ainsi que les données utilisées au cours de notre étude, n'ont évidemment pas permis de construire un modèle permettant de rendre compte de l'évolution récente des prix. Les événements internationaux récents en Irak, au Venezuela ou en Russie ont considérablement perturbés le comportement des indices pétroliers. La méthodologie

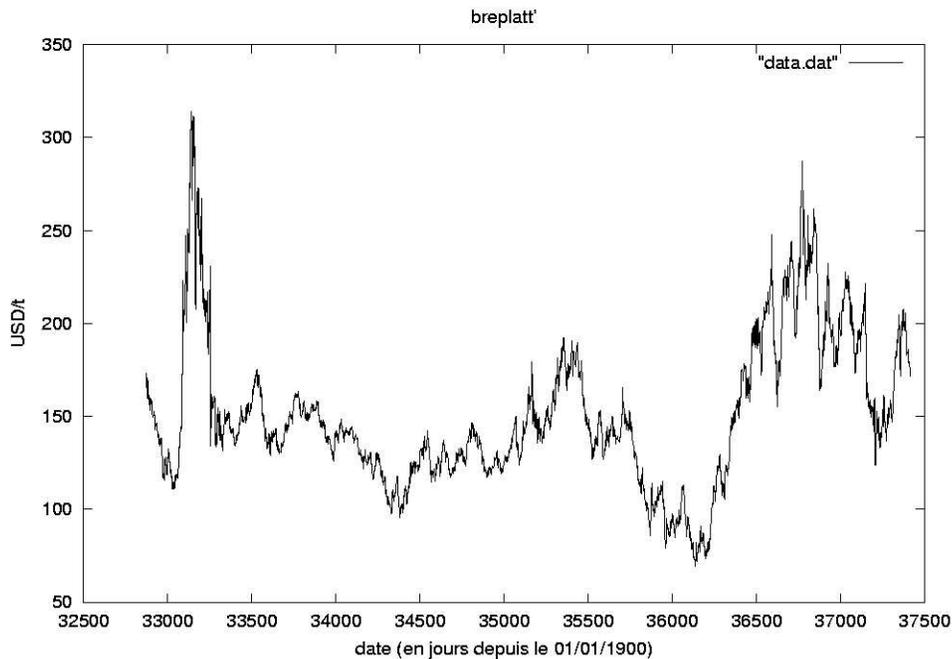


FIG. 1. Evolution caractéristique des indices pétroliers entre 1991 et 2001

que nous allons exposer reste toutefois valide mais les résultats numériques doivent être réactualisés.

2. La construction d'un modèle pour les indices pétroliers

Les données fournies par Gaz de France consistaient en 10 séries de prix journaliers s'étendant sur 10 ans. Le premier objectif de notre étude fut l'élaboration d'un modèle permettant de rendre compte du comportement de ces 10 séries. Deux approches ont été envisagées, la première fondée sur l'étude des séries chronologiques et la seconde sur l'utilisation des processus de diffusion.

2.1. L'approche série chronologique. Les séries chronologiques étaient jusqu'à présent l'outil de modélisation utilisé par Gaz de France. Nous avons donc débuté nos investigations sur cette base. Cette première étude, réalisée à l'aide du logiciel SAS, nous a permis de mettre en évidence la non-stationnarité globale des séries et leur forte corrélation. Par ailleurs, les séries différenciées à l'ordre 1 pouvaient raisonnablement être représentées par un bruit blanc, nous nous sommes donc convaincus qu'un modèle ARIMA(p,1,q) pouvaient convenir. Toutefois, cette approche ne nous a pas donné satisfaction dans la mesure où les algorithmes d'identification de SAS n'ont pour la plupart pas convergé faute de données. Nous nous sommes donc tournés vers les processus de diffusion dont la pratique nous est plus familière.

2.2. L'approche processus de diffusion. Notre objectif devenait donc l'identification, à partir des 10 séries qui nous avaient été fournies, des coefficients de dérive b et de diffusion σ ainsi que la dimension d d'un mouvement brownien W définissant un processus de diffusion X de \mathbb{R}^{10} via l'EDS

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t.$$

2.2.1. *L'étude des corrélations.* Cette première partie avait pour objectif de caractériser rigoureusement les phénomènes de dépendance entre les différentes séries. Ce phénomène intéressait particulièrement Gaz de France, leur permettant d'établir une hiérarchie entre séries directrices et séries subordonnées. Pour notre part, nous y voyions la possibilité de choisir un brownien W de dimension minimale et ainsi de diminuer le coût numérique de la simulation du modèle.

Une étude préliminaire fut menée sur les couples de séries à l'aide du processus de corrélation empirique :

$$\rho_{ij}(t) := \frac{\langle X^i, X^j \rangle_t}{\sqrt{\langle X^i \rangle_t \langle X^j \rangle_t}}.$$

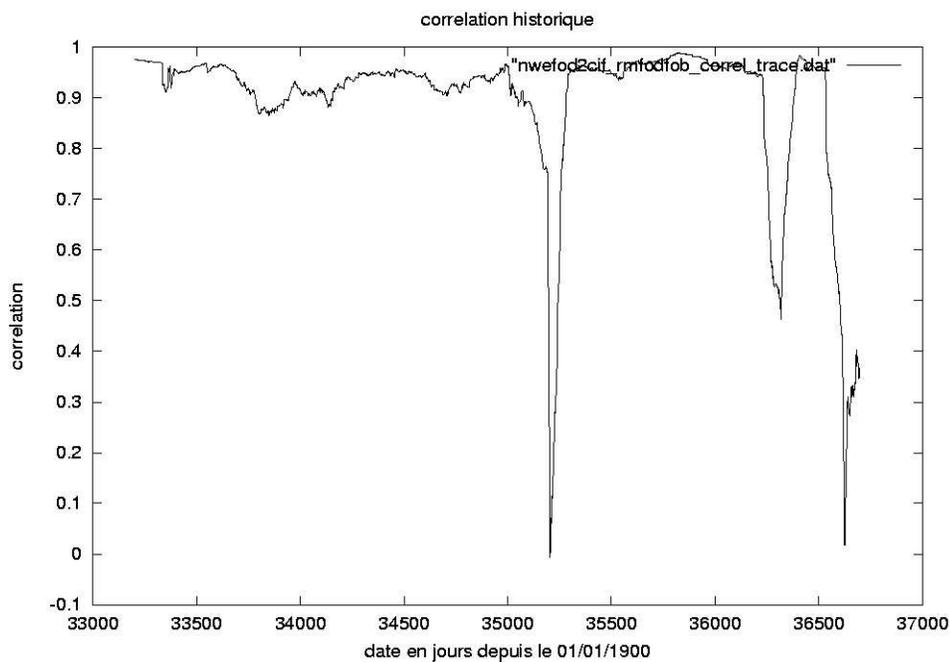


FIG. 2. Exemples de deux séries nettement corrélées

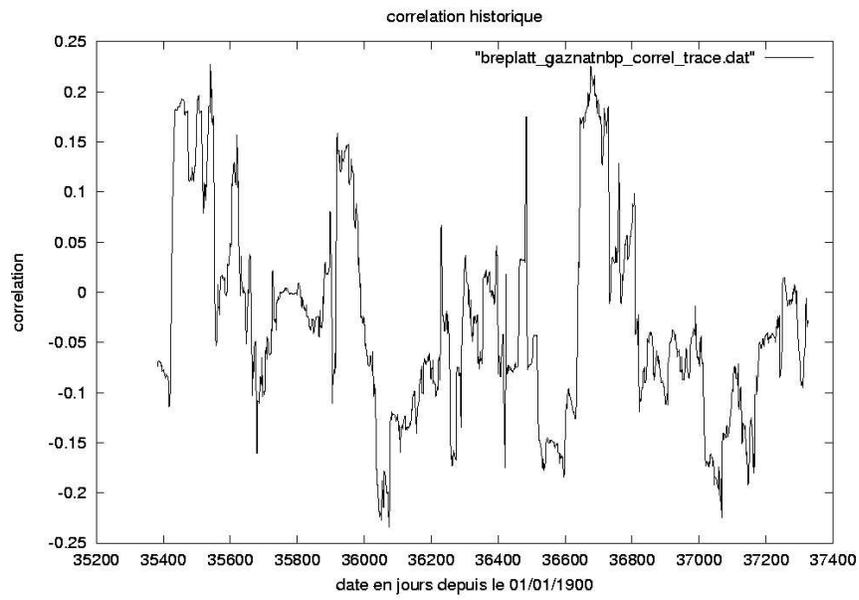


FIG. 3. Deux séries décorrélées

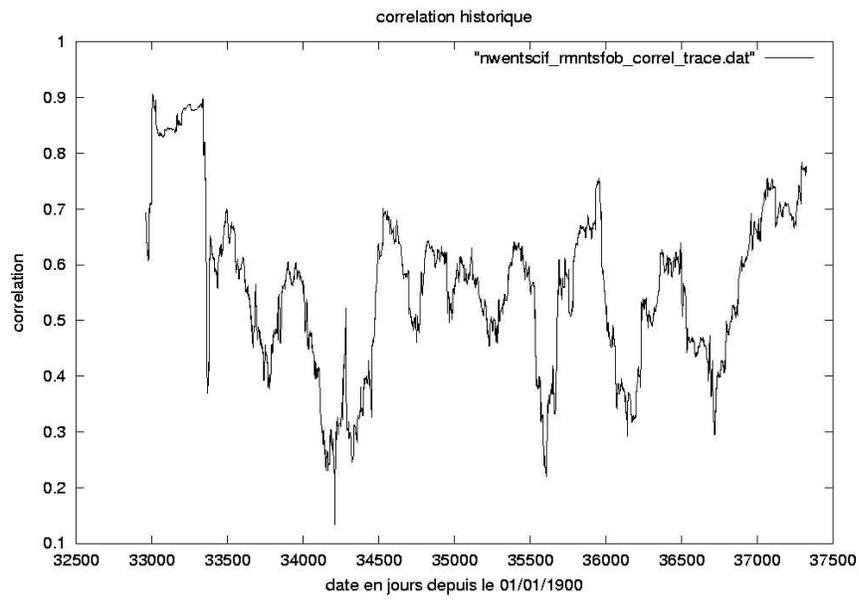


FIG. 4. Deux séries positivement corrélées

Cette étude confirma les résultats déjà observés grâce aux séries chronologiques. Toutefois, pour déterminer la dimension brownienne d et établir des tests d'hypothèses, il nous fallait une approche plus robuste. Celle-ci fut explorée par A. Lejay et L. Lindholm. Leur approche repose sur l'étude des valeurs propres de la matrice de diffusion. Précisément, étant donnée l'observation d'une trajectoire d'un processus d'Itô de \mathbb{R}^n , conduit par un mouvement brownien W de \mathbb{R}^n ,

$$dX_t = b_t dt + \sigma_t dW_t$$

à des instants $t_0, t_1 \dots t_n$, la matrice de diffusion $\sigma_t \sigma_t^*$ peut être estimée localement par la variable

$$\hat{a}_{t_i} := \frac{1}{h} \sum_{j=i-h+1}^i (X_{t_j} - X_{t_{j-1}}) (X_{t_j} - X_{t_{j-1}})^*.$$

Il est ensuite possible de diagonaliser cette matrice et d'estimer ses valeurs propres. Une conséquence du théorème de représentation des martingales permet alors d'affirmer que, si d des ces valeurs propres sont non nulles, alors seules d composantes du mouvement brownien W contribuent à la dynamique du processus X . Néanmoins, en pratique, il est difficile d'établir si une valeur propre est négligeable ou non. Cette procédure nécessiterait donc des résultats la précisant, par exemple au moyen d'intervalles de confiance. Des travaux complémentaires sont poursuivis en ce sens.

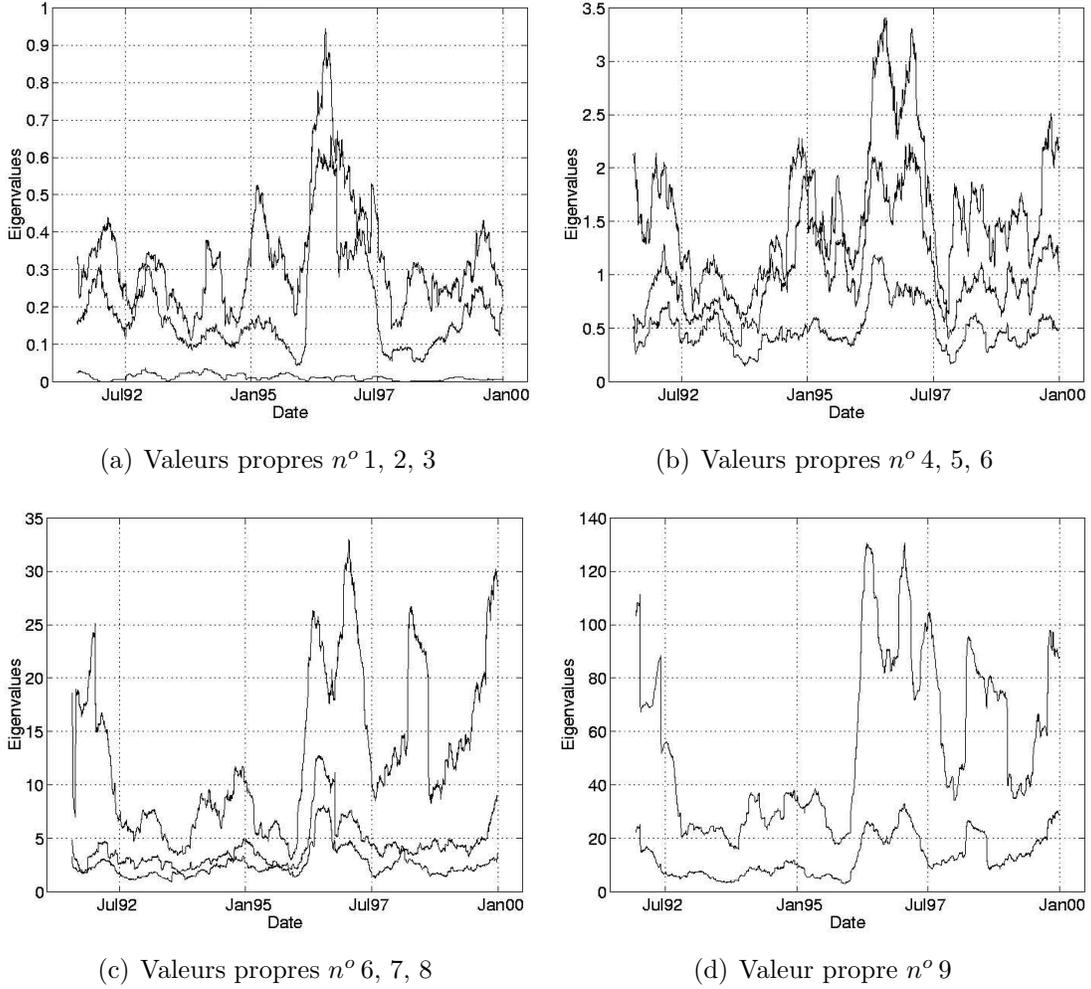


FIG. 5. Estimation des valeurs propres de la matrice de diffusion historique pour les indices pétroliers

2.2.2. *Estimation non-paramétrique du coefficient de diffusion.* Pour estimer le coefficient de diffusion σ , nous avons fait appel à l'estimateur non-paramétrique développé par D. Florens-Zmirou ([30] et [31]) en dimension 1 puis étendu en dimension arbitraire par P. Brugière ([18] et [19]). Mentionnons également les travaux précurseurs de G. Banon [6] et les résultats récents de F. M. Bandi et G. Moloche [5] sur le même thème.

L'estimateur de la fonction de diffusion $a := \sigma\sigma^*$ au point $x \in \mathbb{R}^{10}$ est le suivant :

$$\hat{a}_{n,T}(x) := \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) (X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}) (X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^*}{\sum_{i=1}^{n-1} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)}.$$

T représente la durée des observations disponibles, n , le nombre de données et $\Delta_{n,T} := \frac{T}{n}$. C'est donc un estimateur à noyau fondé sur la variation quadratique empirique. Sous des hypothèses très générales, cet estimateur est fortement consistant et asymptotiquement

normal. Chez D. Florens-Zmirou ([30] et [31]) et P. Brugière ([18] et [19]), le noyau \mathbf{K} est la fonction indicatrice. Mais en pratique, plus que le choix du noyau, c'est la taille de la fenêtre $h_{n,T}$ qui va conditionner la performance de l'estimateur. En théorie, il suffit que cette fenêtre converge à une certaine vitesse vers 0 quand le nombre de données n augmente pour garantir le bon comportement de l'estimateur. Dans le cas de l'estimateur de Florens-Zmirou, nous sommes par exemple conduit à choisir $h_{n,T} = O(n^{-1/\alpha})$ avec $\alpha \in (2, 3)$. Cependant nous ne disposons que d'un nombre de données fixe. Pour choisir convenablement la taille de fenêtre la mieux adaptée nous avons donc eu recours à une procédure de validation croisée (voir par exemple Lo, Mamaysky et Wang [65] pour une introduction et des applications en finance).

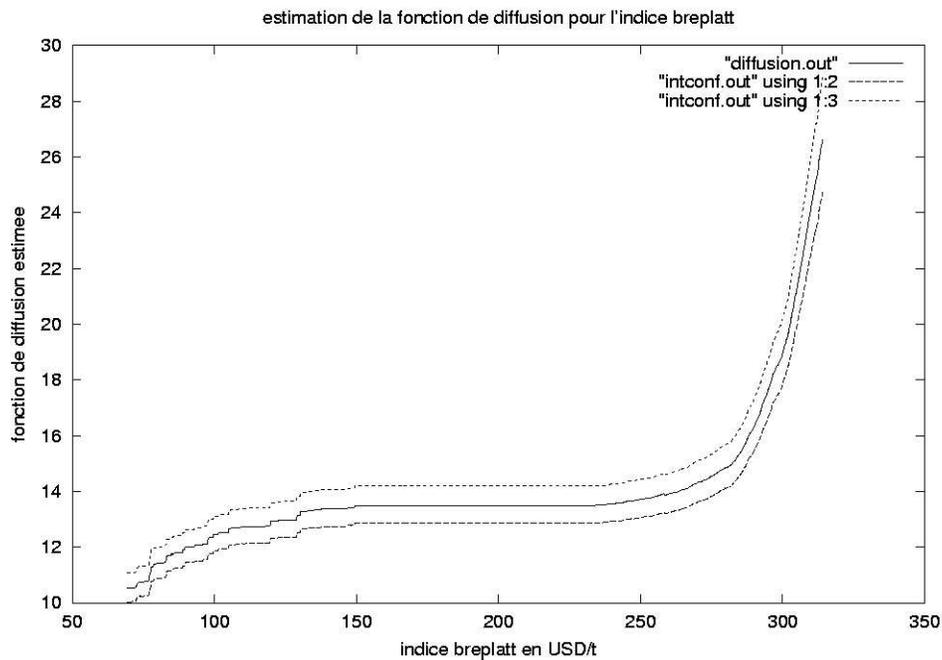
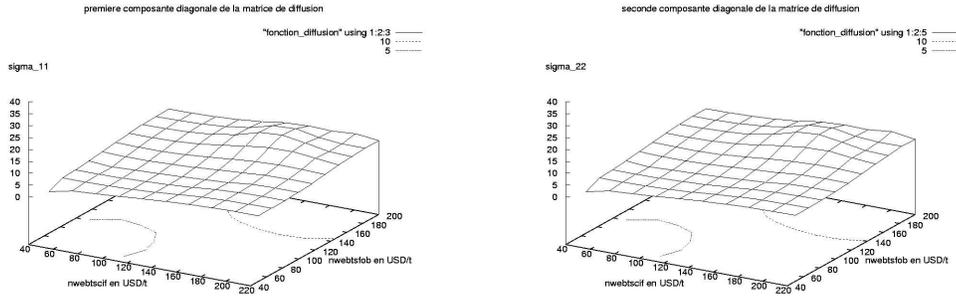
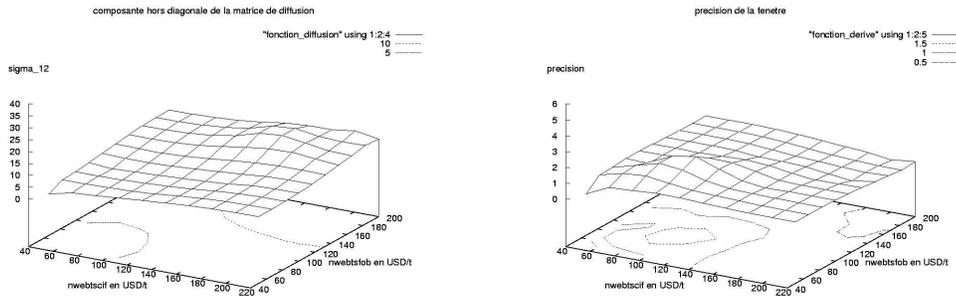


FIG. 6. Estimation unidimensionnelle du coefficient de diffusion et son intervalle de confiance asymptotique à 95%



(a) Première composante diagonale de la matrice de diffusion (b) Seconde composante diagonale de la matrice de diffusion



(c) Composante hors diagonale de la matrice de diffusion (d) Precision de la fenêtre

FIG. 7. Estimation non-paramétrique du coefficient de diffusion pour deux séries

2.2.3. *Estimation paramétrique du coefficient de dérive.* Une fois le coefficient de diffusion caractérisé, il nous restait à estimer le coefficient de dérive. Pour ce faire, nous avons opté pour une approche paramétrique. Le peu de données disponibles ne permettait pas d’espérer de bons résultats dans un cadre non-paramétrique. Néanmoins, les différentes séries laissaient clairement apparaître des stationnarités locales. Nous avons donc retenu un drift de type “retour vers la moyenne” $b(x) = \nu(\theta - x)$ dont les paramètres sont faciles à estimer par une méthode de maximum de vraisemblance (voir le chapitre 3 et Fournié [32]), une fois le coefficient de diffusion connu.

2.2.4. *Synthèse.* La fonctionnelle de prix qui nous a été fournie par Gaz de France pour nos tests numériques fait intervenir les moyennes mobiles de trois indices pétroliers et le taux d’intérêt euro-dollar. Nous avons donc retenu le modèle de diffusion suivant :

$$dS_t = b(S_t) dt + \sigma(S_t) dW_t$$

où S est un processus à valeurs dans \mathbb{R}^n et W est un mouvement brownien standard sur \mathbb{R}^d . Spécifions maintenant les différents paramètres de ce modèle.

Dimensions : $n = 4$ et $d = 4$

Dérive : $b(s)_i = \nu_i (\mu_i - s_i)$, $i = 1 \dots 4$

Diffusion :

$$\sigma(s) = \begin{pmatrix} \sigma_1(s) & 0 & 0 & 0 \\ 0 & \sigma_2(s) & 0 & 0 \\ 0 & 0 & \sigma_3(s) & 0 \\ 0 & 0 & 0 & \sigma_4(s) \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 \\ 0 & \varrho \\ 0 \end{pmatrix}$$

avec

- $\sigma_1(s) = \alpha_1 \sqrt{s_1}$ (Le taux suit le modèle de CIR et est indépendant des trois indices.)

- $\sigma_i(s) = \alpha_i (s_i - K_i)_+ + \beta_i$, $i = 2 \dots 4$

- $\varrho = (\varrho_{kl})_{1 \leq k, l \leq 3}$ telle que $\sum_{k=1}^3 \varrho_{kl}^2 = 1$, $l = 1 \dots 3$

Les valeurs numériques des différents paramètres de ce modèle ont été estimées à partir des résultats obtenus au cours de l'étude décrite précédemment. Nous avons choisi de modéliser le taux d'intérêt euros/dollars par un processus de CIR que nous supposons indépendants des trois autres composantes du processus de diffusion. Ce choix est arbitraire. Il est bien sûr possible d'étudier les corrélations entre ce taux et les indices pétroliers avec les outils que nous avons décrits.

Les figures 8, 9, 10 et 11 présentent une réalisation du modèle que nous venons de présenter. La figure 12 illustre l'évolution du prix simulé à partir de cette réalisation.

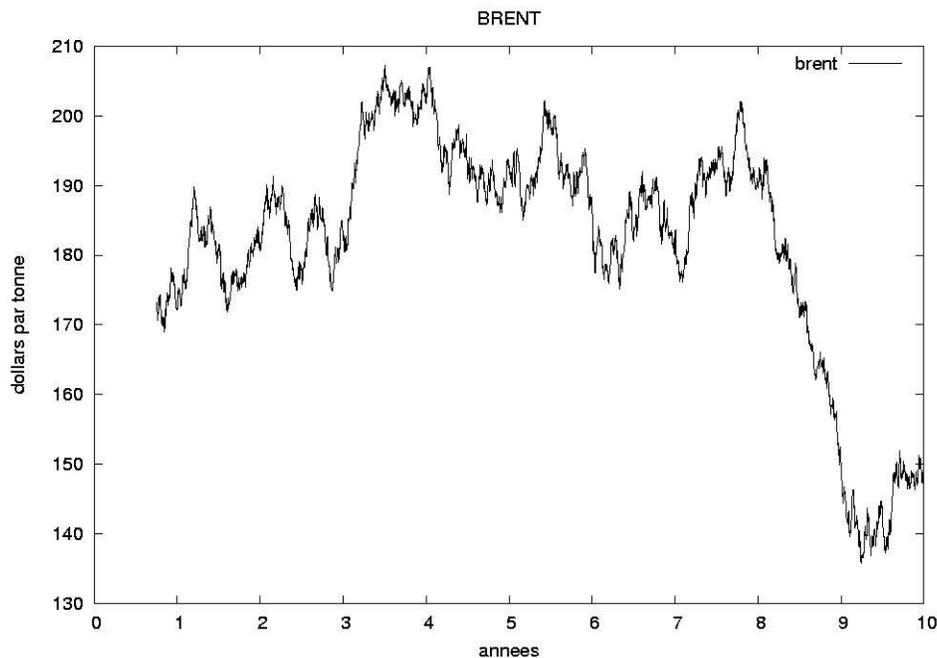


FIG. 8. Cours simulé du Brent

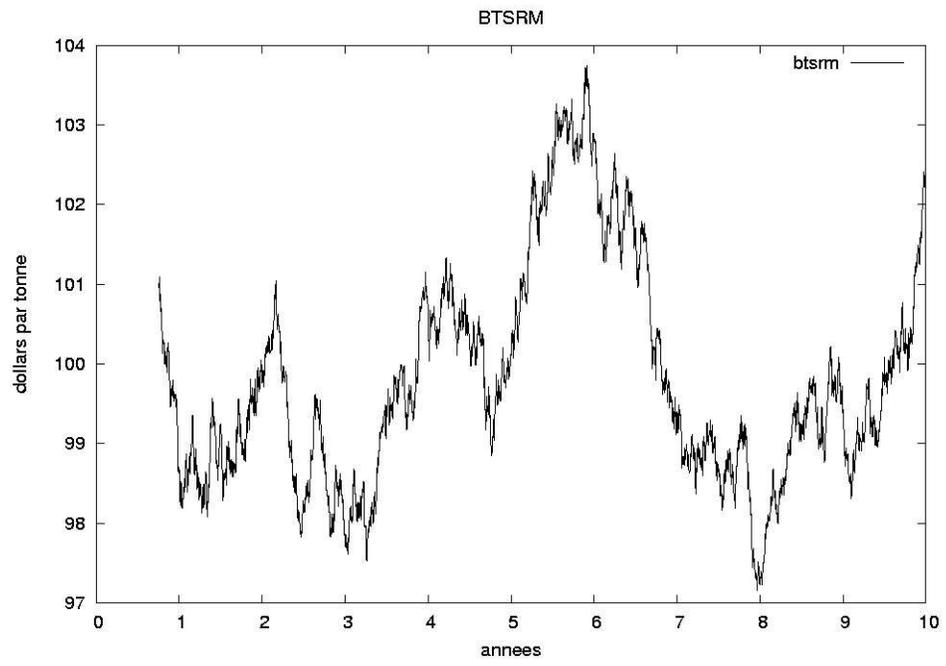


FIG. 9. Cours simulé de Btsrm

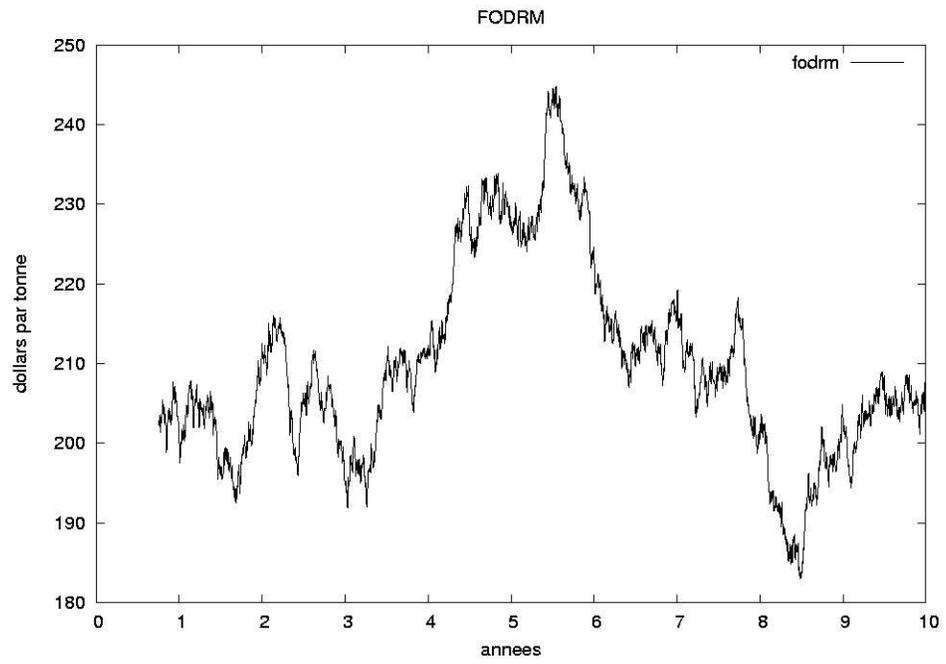


FIG. 10. Cours simulé de Fodrm

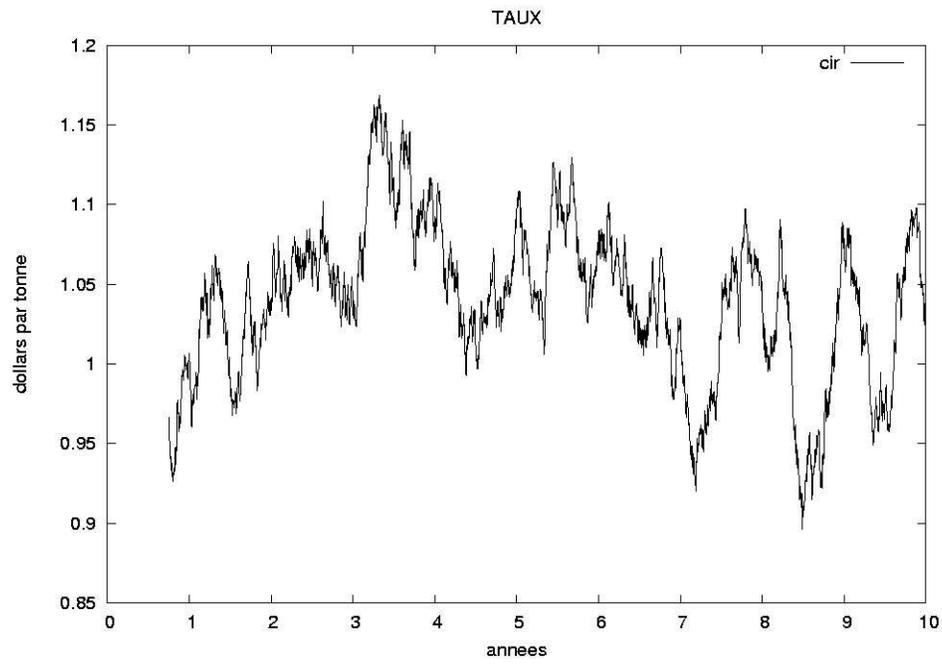


FIG. 11. Cours simulé du taux de change

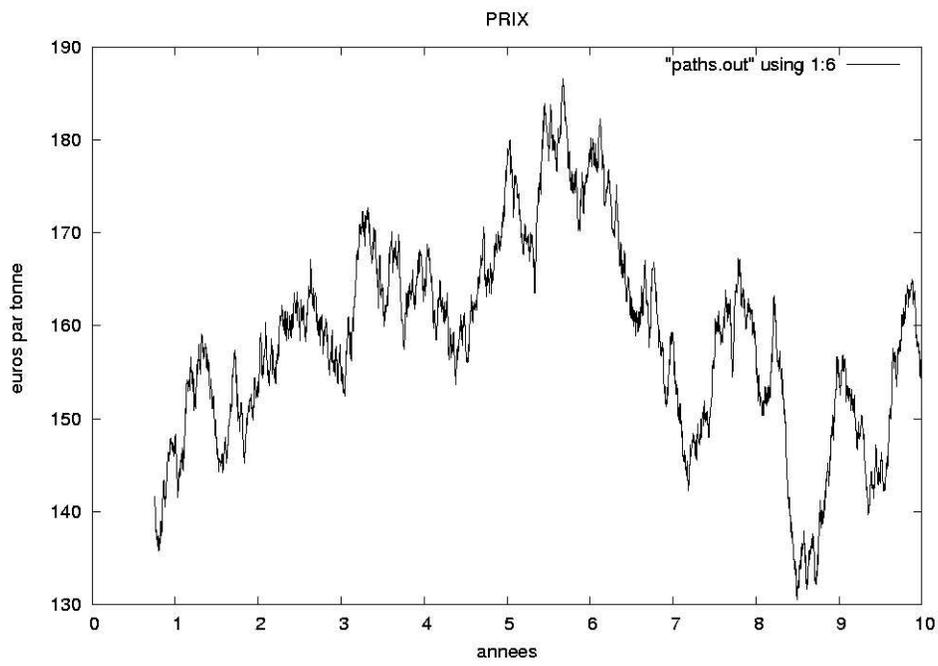


FIG. 12. Evolution simulée du prix

3. La simulation numérique du modèle

La seconde étape de cette collaboration concernait la simulation efficace d'une fonction de prix. Cette fonction prend en argument les séries pétrolières citées précédemment et établit un prix pour le gaz naturel. La transformation se caractérise par une très grande non linéarité. La société Gaz de France était particulièrement intéressée par la simulation de la loi de cette fonctionnelle. La méthode de Monte-Carlo s'imposait donc mais la complexité de la fonctionnelle à simuler laissait présager une variance importante, ce que confirmèrent quelques tests préliminaires. Par ailleurs, les techniques de réduction de variance classiques pouvaient difficilement être mises en œuvre, toujours du fait de l'expression difficilement manipulable de la fonctionnelle de prix. Nous avons donc décidé d'appliquer la technique de réduction de variance adaptative décrite au chapitre 1 à ce problème. Cette technique est très générale et requiert très peu d'information sur les fonctionnelles à simuler. De plus, elle conduit théoriquement à une réduction de variance optimale.

Nous illustrons maintenant l'algorithme de réduction de variance adaptatif décrit au premier chapitre sur trois exemples de fonctionnelles du prix. Pour ce faire, nous avons retenu comme critère la taille de l'intervalle de confiance asymptotique symétrique à 95%, c'est-à-dire la quantité

$$2 \frac{1.96\sigma_n}{\sqrt{n}}$$

où n désigne le nombre d'itérations de l'algorithme de Monte-Carlo et σ_n , l'écart-type estimé à partir de ces n itérations. La figure 13 illustre les résultats sur l'exemple du calcul de la moyenne du prix. Les figures 14 et 15 représentent l'évolution de la taille de l'intervalle de confiance quand la fonctionnelle à estimer est un call sur ce prix.

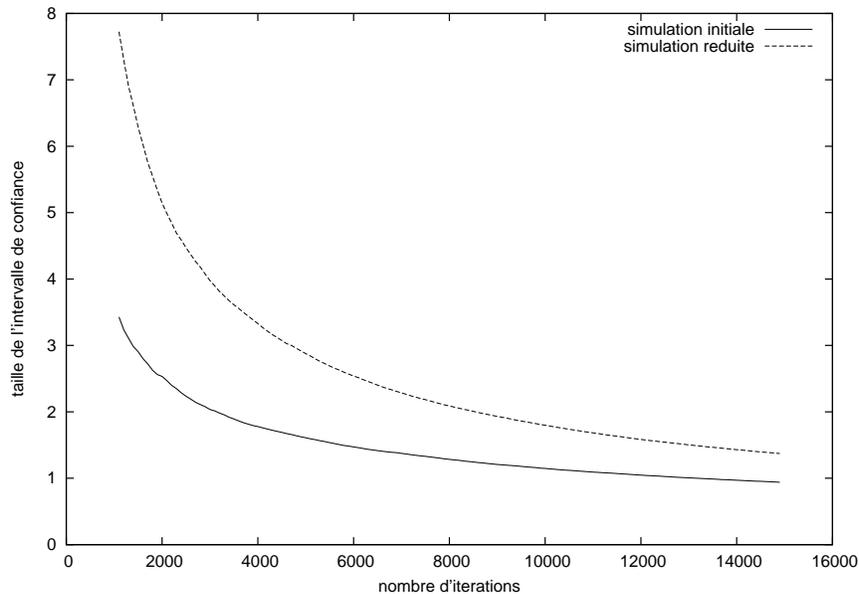


FIG. 13. Evolution de la taille de l'intervalle de confiance pour l'estimation de la moyenne du prix à 10 ans

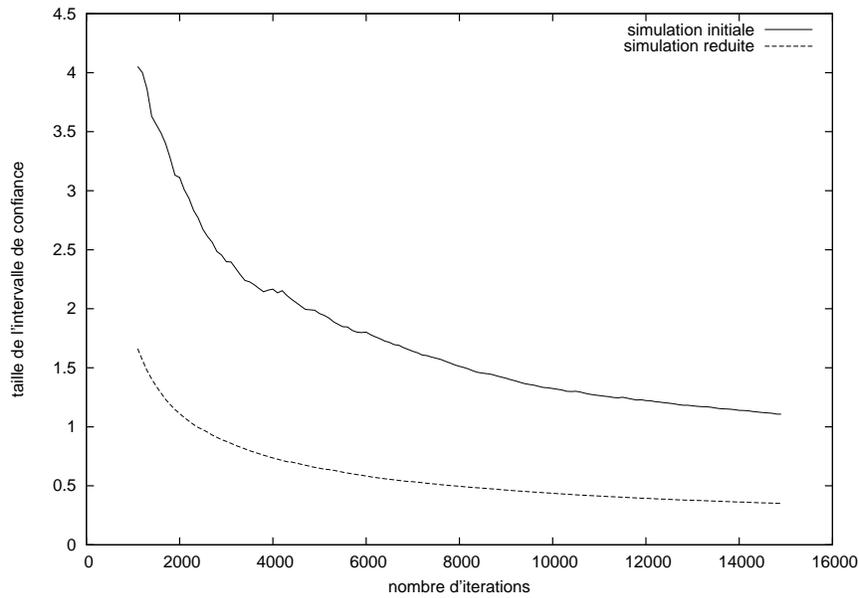


FIG. 14. Evolution de la taille de l'intervalle de confiance pour un call de maturité 10 ans et de strike 150

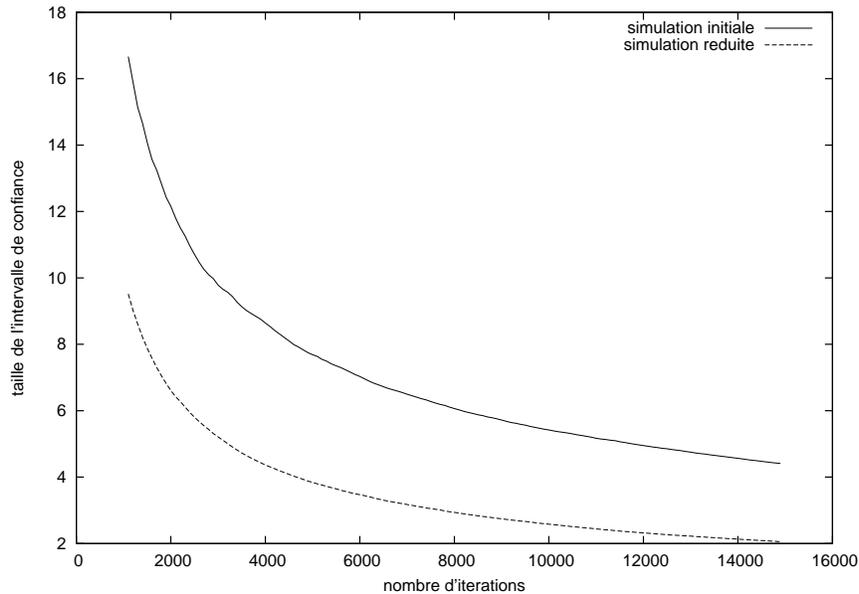


FIG. 15. Evolution de la taille de l'intervalle de confiance pour un actualisé de maturité 10 ans et de strike 160

Notons que sur chacun de ces trois exemples, l'écart-type est au moins divisé par deux. Notre méthode de réduction de variance donne donc des résultats très satisfaisants dans ce contexte. Pour terminer, la figure 16 illustre l'influence du paramètre d'apprentissage sur la réduction de variance effective dans le cas d'un call de maturité 10 ans et de strike 160. Il est clair que choisir une valeur trop petite pour ce paramètre compromet la performance de

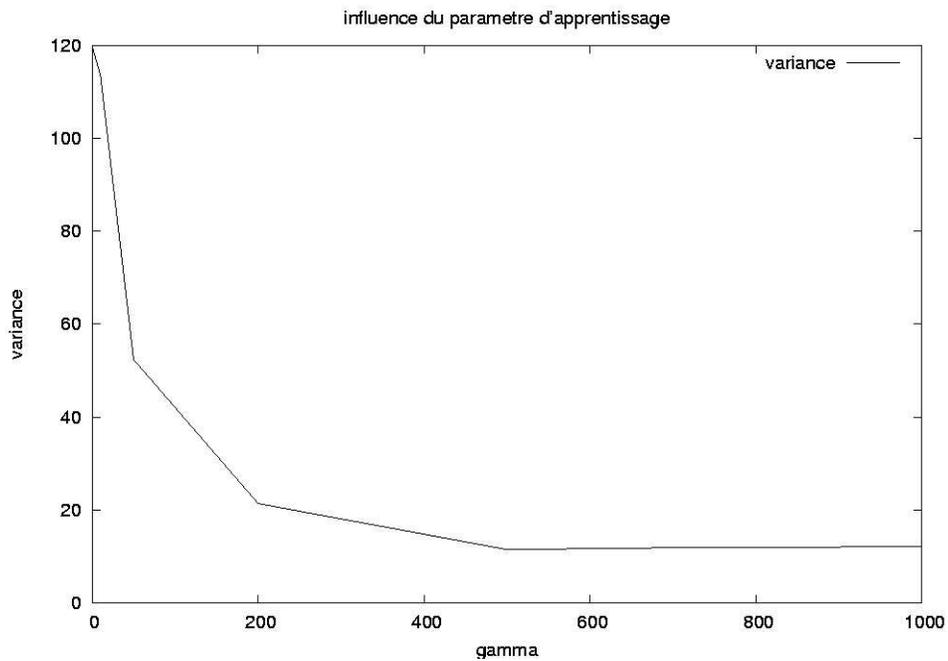


FIG. 16. Influence du paramètre d'apprentissage sur la variance estimée après 16000 itérations Monte-Carlo dans le cas d'un call de maturité 10 ans et de strike 160

l'algorithme. Néanmoins, comme nous l'avons déjà souligné au cours du premier chapitre, des valeurs trop grandes compromettent quant à elles la stabilité de l'algorithme. Le choix du paramètre d'apprentissage est donc essentiel mais délicat. A notre connaissance, il n'existe que très peu de résultats théoriques sur ce sujet.

Bibliographie

1. B. AROUNA, *Robbins-Monro algorithms and variance reduction*, Bachelier Finance Society second world congress, 2002.
2. ———, *Adaptative Monte Carlo method, a variance reduction technique*, Monte Carlo Methods Appl. **10** (2004), no. 1, 1–24.
3. ———, *Robbins–Monro algorithms and variance reduction in finance*, The Journal of Computational Finance **7** (Winter 2003/04), no. 2.
4. D. BABUSIAUX, *Eléments pour l'analyse des évolutions des prix du brut*, Cahiers de l'économie, Analyse et synthèse, vol. 42, Institut Français du pétrole, décembre 2000.
5. F. M. BANDI and G. MOLOCHE, *On the functional estimation of multivariate diffusion processes*, 2001.
6. G. BANON, *Nonparametric Identification for Diffusion Processes*, SIAM Journal of Control and Optimization **16** (1978), no. 3, 380–395.
7. M. BARLOW, K. BURDZY, H. KASPI, and A. MANDELBAUM, *Variably skewed Brownian motion*, Electron. Comm. Probab. **5** (2000), 57–66 (electronic).
8. I. BASAWA and B. RAO, *Statistical Inference for Stochastic Processes*, Academic Press, 1980.
9. M. BASSEVILLE, *Information : entropies, divergences et moyennes*, Tech. Report 1020, IRISA, may 1996, <http://www.irisa.fr/bibli/publi/pi/1996/1020/1020.html>.
10. B. BERCU, *On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications*, Stochastic Process. Appl. **111** (2004), no. 1, 157–173.
11. B. BERCU and A. ROUAULT, *Sharp large deviations for the Ornstein-Uhlenbeck process*, Teor. Veroyatnost. i Primenen. **46** (2001), no. 1, 74–93.
12. P. BERTAIL and S. CLÉMENCON, *Edgeworth expansions of suitably normalized sample mean statistics for atomic markov chains*, Probab. Theory Related Fields (2004).
13. R. N. BHATTACHARYA, *On the functional central limit theorem and the law of the iterated logarithm for Markov processes*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **60** (1982), 185–201.
14. J. P. N. BISHWAL and A. BOSE, *Speed of convergence of the maximum likelihood estimator in the Ornstein-Uhlenbeck process*, Calcutta Statist. Assoc. Bull. **45** (1995), no. 179-180, 245–251.
15. D. BOSQ, *Nonparametric statistics for stochastic processes*, Lecture Notes in Statistics, vol. 110, Springer-Verlag, New York, 1996, Estimation and prediction.
16. D. BOSQ and Yu. DAVYDOV, *Local time and density estimation in continuous time*, Math. Methods Stat. **8** (1999), no. 1, 22–45.
17. P. BOYLE, M. BROADIE, and P. GLASSERMAN, *Monte Carlo methods for security pricing*, Option pricing, interest rates and risk management, Handb. Math. Finance, Cambridge Univ. Press, Cambridge, 2001, pp. 185–238.
18. P. BRUGIÈRE, *Estimation de la variance d'un processus de diffusion dans le cas multidimensionnel*, C. R. Acad. Sci. Paris **t. 312** (1991), no. Série I, 999–1004.
19. ———, *Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus de diffusion multidimensionnelle*, Ann. Inst. Henri Poincaré **29** (1993), no. 3, 357–389.
20. M. E. CABALLERO, B. FERNÁNDEZ, and D. NUALART, *Estimation of densities and applications*, J. Theoret. Probab. **11** (1998), no. 3, 831–851.

21. J. V. CASTELLANA and M. R. LEADBETTER, *On smoothed probability density estimation for stationary processes*, Stochastic Process. Appl. **21** (1986), no. 2, 179–193.
22. F. CHAABANE, *Invariance principles with logarithmic averaging for continuous local martingales*, Statist. Probab. Lett. **59** (2002), no. 2, 209–217.
23. E. CLÉMENT, *Correction du biais pour l'estimation d'une diffusion fondée sur une approximation de la vraisemblance*, C. R. Acad. Sci. Paris Sér. I Math. **322** (1996), no. 3, 281–284.
24. T. G. CONLEY, L. P. HANSEN, and W.-F. LIU, *Bootstrapping the Long Run*, Macroeconomic Dynamics **1** (1997), 279–311.
25. A. DALALYAN, *Estimation non-paramétrique asymptotiquement efficace pour des processus de diffusion ergodiques*, Ph.D. thesis, Université du Maine, 2001.
26. M. DUFLO, *Random iterative models*, Applications of Mathematics (New York), vol. 34, Springer-Verlag, Berlin, 1997, Translated from the 1990 French original by Stephen S. Wilson and revised by the author.
27. B. EFRON, *Bootstrap methods : another look at the jackknife*, Ann. Statist. **7** (1979), no. 1, 1–26.
28. W. FELLER, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971.
29. D. FLORENS-LANDAIS and H. PHAM, *Large deviations in estimation of an Ornstein-Uhlenbeck model*, J. Appl. Probab. **36** (1999), no. 1, 60–77.
30. D. FLORENS-ZMIROU, *Estimation de la variance d'une diffusion à partir d'une observation discrétisée*, Note au Compte-Rendu de l'Académie des Sciences **309** (1989), no. I, 195–200.
31. ———, *On estimating the diffusion coefficient from discrete observations*, J. Appl. Prob. **30** (1993), 790–804.
32. E. FOURNIÉ, *Statistiques des Diffusions Ergodiques avec Applications en Finance*, Ph.D. thesis, Université de Nice-Sophia Antipolis, 1993.
33. E. FOURNIÉ, J.-M. LASRY, J. LEBUCHOUX, and P.-L. LIONS, *Applications of Malliavin calculus to Monte-Carlo methods in finance. ii*, Finance and Stochastics **5** (2001), 201–236.
34. E. FOURNIÉ, J.-M. LASRY, J. LEBUCHOUX, P.-L. LIONS, and N. TOUZI, *Applications of Malliavin calculus to Monte-Carlo methods in finance*, Finance and Stochastics **3** (1999), 391–412.
35. H. GANIDIS, B. ROYNETTE, and F. SIMONOT, *Convergence rate of some semi-groups to their invariant probability*, Stochastic Process. Appl. **79** (1999), no. 2, 243–263.
36. H. GEMAN and M. YOR, *Bessel processes, Asian options, and perpetuities.*, Math. Finance **3** (1993), no. 4, 349–375.
37. Ī. Ī. GIĪHMAN and A. V. SKOROHOD, *Stochastic Differential Equations*, Springer-Verlag, New York, 1972, Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72.
38. P. GLASSERMAN, P. HEIDELBERGER, and P. SHAHABUDDIN, *Asymptotically optimal importance sampling and stratification for pricing path-dependent options*, Math. Finance **9** (1999), no. 2, 117–152.
39. E. GOBET and A. KOHATSU-HIGA, *Computation of Greeks for Barrier and Lookback Options using Malliavin Calculus*, Electronic Communications in Probability (2003), no. 8.
40. C. GOURIEROUX, A. MONFORT, and E. RENAULT, *Indirect Inference*, Journal of Applied Econometrics (1993), no. 8, 85–118.
41. P. HALL, *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, Springer-Verlag, New York, 1992.
42. R. Z. HAS'MINSKII, *Stochastic stability of differential equations*, Monographs and Textbooks on Mechanics of Solids and Fluids : Mechanics and Analysis, vol. 7, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980, Translated from the Russian by D. Louvish.
43. C. HESS, *Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator*, Ann. Statist. **24** (1996), no. 3.

44. N. IKEDA and S. WATANABE, *Stochastic Differential Equations and Diffusion Processes (2nd Ed.)*, North Holland, 1989.
45. K. ITÔ and H. P. McKEAN Jr., *Diffusion processes and their sample paths*, Springer-Verlag, Berlin, 1974, Second printing, corrected, Die Grundlehren der mathematischen Wissenschaften, Band 125.
46. Y. KAKIZAWA, *Edgeworth approximation in the O-U process*, Sankhyā : The Indian Journal of Statistics **64** (2000).
47. I. KARATZAS and S.E. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.
48. A. KOHATSU-HIGA and M. MONTERO, *Malliavin Calculus in Finance*, To appear in Handbook of Computational Finance. Birkhauser, 2003.
49. A. KOHATSU-HIGA and R. PETTERSON, *Variance reduction methods for simulation of densities on Wiener space*, SIAM Journal of Numerical Analysis (2002).
50. I. KONTOYIANNIS and S. P. MEYN, *Spectral theory and limit theorems for geometrically ergodic Markov processes*, Ann. Appl. Probab. **13** (2003), no. 1, 304–362.
51. N. V. KRYLOV, *Controlled diffusion processes*, Applications of Mathematics, vol. 14, Springer-Verlag, New York, 1980, Translated from the Russian by A. B. Aries.
52. H. KUNITA, *Stochastic differential equations and stochastic flows of diffeomorphisms*, Cours de l'Ecole d'été de Probabilité de Saint-Flour 1982 (Springer, ed.), Lecture Notes in Mathematics, 1984.
53. H. J. KUSHNER and G. G. YIN, *Stochastic approximation and recursive algorithms and applications*, second ed., Applications of Mathematics (New York), vol. 35, Springer-Verlag, New York, 2003, Stochastic Modelling and Applied Probability.
54. Y. A. KUTOYANTS, *On a hypotheses testing problem and asymptotic normality of stochastic integral*, Theory of probability and its applications (1975).
55. Y.A. KUTOYANTS, *Parameter Estimation for Stochastic Processes*, Heldermann Verlag, 1984.
56. Yu. A. KUTOYANTS, *Statistical inference for ergodic diffusion processes.*, Springer Series in Statistics, Springer-Verlag, 2003.
57. O. A. LADYŽENSKAJA, V. A. SOLONNIKOV, and N. N. URAL'CEVA, *Linear and quasilinear equations of parabolic type*, Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23, American Mathematical Society, Providence, R.I., 1967.
58. D. LAMBERTON and G. PAGÈS, *Recursive computation of the invariant distribution of a diffusion*, Bernoulli **8** (2002), no. 3, 367–405.
59. ———, *Recursive computation of the invariant distribution of a diffusion : the case of a weakly mean reverting drift*, Stoch. Dyn. **3** (2003), no. 4, 435–451.
60. P. LEZAUD, *Chernoff and Berry-Esséen inequalities for Markov processes*, ESAIM Probab. Statist. **5** (2001), 183–201 (electronic).
61. R. LIPTSER, *Large deviations for two scaled diffusions*, Probab. Theory Related Fields **106** (1996), no. 1, 71–104.
62. R. LIPTSER and V. SPOKOINY, *Moderate deviations type evaluation for integral functionals of diffusion processes*, Electron. J. Probab. **4** (1999), no. 17, 25 pp. (electronic).
63. R. LIPTSER, V. SPOKOINY, and A. Yu. VERETENNIKOV, *Freidlin-Wentzell type large deviations for smooth processes*, Markov Process. Related Fields **8** (2002), no. 4, 611–636.
64. R.S. LIPTSER and A.N. SHIRYAYEV, *Statistics of Random Processes I, II*, Springer-Verlag, 1978.
65. A. W. LO, H. MAMAYSKY, and J. WANG, *Foundations of Technical Analysis : Computational Algorithms, Statistical Inference, and Empirical Implementation*, The Journal of Finance (2000).
66. F. MAAOUIA, *Principes d'invariance par moyennisation logarithmique pour les processus de Markov*, Ann. Probab. **29** (2001), no. 4, 1859–1902.

67. M. MARTINEZ, *Interprétations probabilistes d'opérateurs sous forme divergence et analyse de méthodes numériques probabilistes associées*, Ph.D. thesis, Université de Marseilles, 2004.
68. J. C. MATTINGLY, A. M. STUART, and D. J. HIGHAM, *Ergodicity for SDEs and approximations : locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl. **101** (2002), no. 2, 185–232.
69. S.P. MEYN and R.L. TWEEDIE, *Markov Chains and Stochastic Stability*, Springer-Verlag, 1993.
70. P. A. MYKLAND, *Embedding and asymptotic expansions for martingales*, Probab. Theory Related Fields **103** (1995), no. 4, 475–492.
71. N.J. NEWTON, *Variance reduction for simulated diffusions*, SIAM J. Appl. Math. **54** (1994), no. 6, 1780–1805.
72. E. PARDOUX and A. Yu. VERETENNIKOV, *On the Poisson equation and diffusion approximation. I*, Ann. Probab. **29** (2001), no. 3.
73. V.V. PETROV, *Sums of Independent Random Variables*, Springer-Verlag, 1975.
74. P. C. B. PHILLIPS and J. YU, *Jackknifing Bond Option Prices*, Tech. report, Cowles Foundation for Research in Economics, Yale University, 2003.
75. P. PROTTER, *Stochastic Integration and Differential Equations*, Springer-Verlag, Berlin, 1990.
76. R. REIDER, *An Efficient Monte Carlo Technique for Pricing Options*, Tech. report, Wharton School, University of Pennsylvania, 1993.
77. D. REVUZ and M. YOR, *Continuous Martingales and Brownian Motion*, Springer-Verlag, Berlin, 1991.
78. L. C. G. ROGERS and D. WILLIAMS, *Diffusions, Markov processes, and Martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition.
79. Y. SAKAMOTO and N. YOSHIDA, *Asymptotic expansion of M -estimator over Wiener space*, Stat. Inference Stoch. Process. **1** (1998), no. 1, 85–103.
80. C. SOIZE, *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, Series on Advances in Mathematics for Applied Sciences, vol. 17, World Scientific Publishing Co. Inc., River Edge, NJ, 1994.
81. D. W. STROOCK and S. R. S. VARADHAN, *Multidimensional diffusion processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 233, Springer-Verlag, Berlin, 1979.
82. Y. SU and M. C. FU, *Optimal importance sampling in securities pricing*, Journal of Computational Finance **5** (2002), no. 4, 27–50.
83. D. TALAY, *Second-order discretization schemes of stochastic differential systems for the computation of the invariant law.*, Stochastics Stochastics Rep. **29** (1990), no. 1, 13–36.
84. ———, *Stochastic Hamiltonian systems : exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Process. Related Fields **8** (2002), no. 2, 163–198, Inhomogeneous random systems (Cergy-Pontoise, 2001).
85. D. TALAY and L. TUBARO, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stoch. Anal. Appl. **8** (1990), no. 4, 94–120.
86. F. J. VASQUEZ-ABAD and D. DUFRESNE, *Accelerated Simulation for Pricing Asian Options*, Proceedings of the 1998 Winter Simulation Conference (D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, eds.), 1998.
87. N. YOSHIDA, *Malliavin calculus and asymptotic expansion for martingales*, Probab. Theory Related Fields **109** (1997), no. 3, 301–342.
88. J. H. VAN ZANTEN, *On the uniform convergence of the empirical density of an ergodic diffusion*, Stat. Inference Stoch. Process. **3** (2000), no. 3, 251–262 (2001).

Résumé : Les propriétés asymptotiques des algorithmes de type Monte-Carlo et des fonctionnelles usuelles de processus de diffusions ergodiques se caractérisent à l'aide de théorèmes de la limite centrale. L'objet de cette thèse est la présentation de résultats raffinant ces théorèmes dans quatre cadres différents.

La première partie de ce travail concerne la simulation des processus de diffusion. Le premier chapitre est consacré à la présentation d'une méthode permettant de contrôler adaptativement la variance au cours d'une simulation Monte-Carlo. Des applications sont données en finance. Le second chapitre propose un estimateur de la variance asymptotique des simulations ergodiques. Sa construction repose sur des résultats de type théorème de la limite centrale presque sûr. Des techniques de réduction de variance sont proposées dans ce cadre. La seconde partie concerne la statistique des processus. Le premier chapitre traite des processus de diffusion ergodiques. Pour différentes fonctionnelles de ces processus, nous démontrons des développements d'Edgeworth précisant la vitesse de convergence du théorème de la limite centrale. Des applications en statistiques sont proposées, et notamment une ouverture vers le bootstrap. Le second chapitre propose un cadre théorique pour l'estimation paramétrique de processus de diffusions généralisant le mouvement brownien asymétrique.

Mots-clés : Diffusions, Monte-Carlo, Réduction de Variance, Développements d'Edgeworth, Algorithme de Robbins-Monro, Ergodicité, Principes d'invariance avec poids logarithmiques, Equation de Poisson, Bootstrap, Diffusions asymétriques

Abstract : Central limit theorems characterize the asymptotic properties of Monte-Carlo algorithms and usual functionals of ergodic diffusion processes. The aim of this thesis is the introduction of results precisizing these theorems in four different areas.

The first part of this work is devoted to the simulation of diffusion processes. The first chapter introduces a numerical method leading to a dynamic control of the variance within a Monte-Carlo algorithm. Applications in finance are provided. The second chapter deals with an efficient estimator for the variance of ergodic simulations. Its properties rely on an almost sure central limit theorem. Variance reduction technics are devised in this framework.

The second part of this work is devoted to the statistical estimation of diffusion processes. The first chapter deals with ergodic processes. For different functionals of such processes, we prove Edgeworth expansions precisizing the speed of convergence in the central limit theorem. Applications in statistics are provided, especially a proof of the efficiency of the Bootstrap in this setting. The second chapter introduces a framework for the parametric estimation of processes generalizing the skew Brownian motion.

Key Words : Diffusions, Monte-Carlo, Variance reduction, Edgeworth expansions, Robbins-Monro algorithm, Ergodicity, Invariance principles with logarithmic averaging, Poisson equation, Bootstrap, Skew processes