UNIVERSITÉ PIERRE ET MARIE CURIE – PARIS 6

T H È S E

pour obtenir le titre de

Docteur de l'Université Pierre et Marie Curie

Spécialité : Mathématiques Appliquées

Présentée et soutenue par

Daphné Giorgi

Théorèmes limites pour estimateurs Multilevel avec et sans poids. Comparaisons et applications.

Thèse dirigée par Gilles Pagès

préparée au LPMA, avec le soutien de la Chaire Risques Financiers de la Fondation du Risque de l'Institut Louis Bachelier

soutenue le 2 juin 2017

Jury :

Rapporteurs :	Mike Giles	-	University of Oxford
	Benjamin Jourdain	-	École des Ponts ParisTech
Directeur :	Gilles Pagès	-	Université Pierre et Marie Curie
Examinateurs :	Nicole El Karoui	-	Université Pierre et Marie Curie
	Ahmed Kebaier	-	Université Paris 13
	Vincent Lemaire	-	Université Pierre et Marie Curie

Remerciements

Je remercie en premier lieu mon directeur de thèse, Gilles Pagès, qui a gentiment accepté, bien que je sois *au milieu du chemin de ma vie*, ce rôle de Virgile, en faisant preuve d'une grande patience tout au long de ce parcours, en ne me refusant jamais une explication, en me donnant du courage et en m'éperonnant au besoin, lui qui a surtout été infatigable et généreux dans tous ces rôles.

Je souhaite en deuxième lieu remercier Vincent Lemaire, qui, dans les faits, a été co-directeur de cette thèse. J'apprécie chez lui son honnêteté et sa rigueur, sa gentillesse et sa sensibilité, pour ne pas parler de son code, qui est le reflet de l'ordre qu'il a dans la tête.

Je remercie ensuite les rapporteurs, Mike Giles et Benjamin Jourdain. Je suis très honorée qu'ils aient accepté cette charge, le contenu de ce manuscrit s'inspire de leur travail, que je respecte énormément, et je n'aurais pas pu espérer avoir rapporteurs mieux qualifiés. Je remercie aussi les autres membres du Jury d'avoir accepté de participer à ma soutenance, Nicole El Karoui, que je remercie pour sa force et son regard, et Ahmed Kebaier, que je suis heureuse d'avoir pu connaître avant la soutenance et que je remercie pour son accueil.

Je passe ensuite aux collègues du LPMA, qui sont pour moi une grande famille. Je remercie d'abord notre directeur Francis de s'être patiemment et régulièrement soucié de mon parcours pendant ces années, suivi par Lorenzo, sans qui je ne serais pas là, et qui en plus d'être un collègue est un ami. Je remercie ensuite Thomas pour sa capacité à tourner en blague tout moment de désespoir et pour plein d'autres bonnes raisons, parmi lesquelles sa passion pour les insectes, Thierry parce que c'est un probabiliste, mais aussi un violoniste, un cycliste et un sempiternel déménageur, Zhan pour sa douceur, ses plantes et les gâteaux chinois, Nathalie, en parlant de plantes, pour nos conversations sans fin sur l'état de santé de notre petit commerce verdoyant et pour son amour de l'hiver et du froid, Catherine pour sa répartie et son humour impitoyable, son vélo et ses embrouilles dans la rue, et cette tendresse qu'elle a parfois du mal à cacher, Tabea pour ses boîtes de biscuits de Noël, de cotons-tiges et de vêtements, Mathieu parce que c'est le plus sympa, Fanny parce qu'elle ne finira jamais de déballer ses cartons, Étienne pour me faire complètement confiance en termes de vélo et de pneus crevés, Ismaël parce qu'il est marié à la femme qui a le plus beau prénom du monde, Quentin pour son coté bricoleur, Damien à qui je souhaite passer l'information que *ryada* en arabe signifie à la fois sport et maths et que du coup il peut dire qu'il est bon en sport, Frédérique pour la photo de sa fille qui lui ressemble comme une goutte d'eau, Omer pour donner de la notoriété au laboratoire, Eric pour sa passion pour le tennis, Cédric parce qu'il attend avec impatience le moment où on aura notre propre ownCloud, Lokmane parce qu'il sait utiliser une carte graphique pour autre chose que des jeux vidéos, Jean Paul pour son curieux invité et Yves pour les discussions sur le cinéma et les séries, Nicolas parce que s'il doit aller à l'École des Ponts il y va à vélo, François parce qu'il envoie Ayman à Roscoff, Michelle pour sa gentillesse, Romain pour avoir ajouté une couche d'humour grinçant aux pauses café, Gregory pour m'avoir initiée au Shiny, Raphael pour regretter son adresse ip fixe, David pour la permaculture, Camille pour ses descriptions des *cesti* envoyés depuis l'Italie, Dan excellent compagnon de bureau, Julien pour m'avoir supporté en Master 2, Serena pour sa patience avec le site web du laboratoire, Florence pour son suivi parmi les méandres de la bureaucratie de la grande tour centrale et

pour l'art du tricot, Josette pour sa gentillesse, sa joie et son rire à nul autre pareil, Aldjia précieuse femme aux mille casquettes, Khash sans qui j'aurais été perdue bien des fois et dont la seule existence suffit à nous donner à tous un sommeil tranquille, Clément pour sa constante et patiente présence et pour sa disponibilité, Altair dont j'ai regretté le départ mais à qui je souhaite de réussir dans tous ses projets, où qu'ils soient, Jacques pour son accueil, ses gentils conseils et son érudition, Philippe pour la psychomagie, Corentin pour ses explications minutieuses et sa patience séraphique, et parmi les doctorants je remercie Sarah pour nos conversations, Olga pour son énergie, Nicolas pour être bien geek sur les bords et Thibaut pour le grand coup de main qu'il nous a donné pour le(s) pot(s) de thèse. J'en oublie sans doute, alors j'espère rattraper les derniers en remerciant tous ceux qui ont espéré que le prochain nom de laboratoire soit La/Ma/Ta ProStat.

Une pensée spéciale va à mes amis, à Nadia et à mes camarades de cours d'égyptien, à Mat' les vacances, alle belle figlie dell'amore, agli zingari e alle zingarate, aux réparateurs de vélos, aux mains vertes, aux Chapoys, aux Parisiens et aux Juvisiens, aux Cachanais, aux joueurs de cartes et de badminton, aux Italiens à Paris, aux fans de Gianna Nannini, à Arsita, alla sagra del daino, au plus cool de nos amis Oluc, au Caire, au camping, et si j'ai oublié quelqu'un à la montagne et à la mer.

Enfin je remercie ma famille. Mon frère pour la persévérance qu'il met dans ses passions depuis qu'il est tout petit, parce qu'il travaille pour nos vacances et parce que ce n'est pas qu'un frère, c'est un ami. Mes parents parce qu'ils me soutiennent depuis toujours, parce que nous n'avons jamais manqué d'amour et que même si j'allais en prison, ils viendraient quand même me ramener des oranges. Ayman, *habib elbi*, parce que la vie à ses côtés déborde de rires et de couleurs et parce qu'il continue de m'aimer, tout en sachant d'où je suis sortie. Et pour finir je te remercie toi, *batatsayya*, petite chose qui n'a pas encore de prénom, mais qui fait déjà notre joie.

Théorèmes limites pour estimateurs Multilevel avec et sans poids. Comparaisons et applications.

Résumé : Dans ce travail, nous nous intéressons aux estimateurs Multilevel Monte Carlo. Ces estimateurs vont apparaître sous leur forme standard, avec des poids et dans une forme randomisée. Nous allons rappeler leurs définitions et les résultats existants concernant ces estimateurs en termes de minimisation du coût de simulation. Nous allons ensuite montrer une loi forte des grands nombres et un théorème central limite. Après cela nous allons étudier deux cadres d'applications. Le premier est celui des diffusions avec schémas de discrétisation antithétiques, où nous allons étendre les estimateurs Multilevel aux estimateurs Multilevel avec poids. Le deuxième est le cadre nested, où nous allons nous concentrer sur les hypothèses d'erreur forte et faible. Nous allons conclure par l'implémentation de la forme randomisée des estimateurs Multilevel, en la comparant aux estimateurs Multilevel avec et sans poids.

Mots clés : Monte Carlo, Multilevel Monte Carlo, Multilevel Monte Carlo avec poids, Multilevel Monte Carlo Randomisé, Loi Forte des Grands Nombres, Théorème Central Limite, Schéma de discrétisation d'Euler, Schéma de discrétisation de Milstein, Schémas de discrétisation Antithetiques, nested Monte Carlo, Extrapolation de Richardson Romberg , Statistical Romberg method, Variable de contrôle, Réduction de variance, Convergence forte, Convergence faible.

Limit theorems for Multilevel estimators with and without weights. Comparisons and applications.

Abstract : In this work, we will focus on the Multilevel Monte Carlo estimators. These estimators will appear in their standard form, with weights and in their randomized form. We will recall the previous existing results concerning these estimators, in terms of minimization of the simulation cost. We will then show a strong law of large numbers and a central limit theorem. After that, we will focus on two application frameworks. The first one is the diffusions framework with antithetic discretization schemes, where we will extend the Multilevel estimators to Multilevel estimators with weights, and the second is the nested framework, where we will analyze the weak and the strong error assumptions. We will conclude by implementing the randomized form of the Multilevel estimators, comparing this to the Multilevel estimators with and without weights.

Keywords : Monte Carlo, Multilevel Monte Carlo, Weighted Multilevel Monte Carlo, Randomized Multilevel Monte Carlo, Strong Law of Large Numbers, Central Limit Theorem, Euler discretization scheme, Milstein discretization scheme, Antithetic discretization schemes, nested Monte Carlo, Richardson Romberg extrapolation, Statistical Romberg method, Control variable, Variance reduction, Strong convergence, Weak convergence.

Table des matières

Table des figures

Liste des tableaux

Introduction

Cette thèse a été effectuée grâce au soutien de la Chaire Risques Financiers de la Fondation du Risque de l'Institut Louis Bachelier. Cette fondation a pour objet de contribuer durablement au développement du potentiel français de recherche, d'enseignement et de formation dans tous les domaines du risque : risques financiers, risques industriels, risques environnementaux, risques patrimoniaux, risques de santé, etc. Depuis sa création en 2007, la Fondation du Risque s'attache à initier des actions de recherche et de développement, notamment par la promotion de projets d'enseignement et de recherche et la diffusion des connaissances à travers des programmes de formation de haut niveau. Elle a également vocation à favoriser la pédagogie du risque par la diffusion de connaissances auprès du grand public comme d'un public averti.

Dans ce manuscrit nous donnons le détail des résultats de recherche qui ont été obtenus dans le cadre d'un contrat d'Ingénieur de Recherche et Développement d'une durée de trois ans. Pour ce qui concerne la partie Développement du contrat, une interface web permettant de calculer des estimateurs Multilevel a été mise en œuvre, afin de mettre les résultats à disposition d'un utilisateur générique et en prévision du développement de relations industrielles. L'application web est écrite en Python, en utilisant le module Flask, avec l'ajout de HTML et JavaScript. Le calcul des paramètres et des estimateurs eux-mêmes est fait à travers l'appel du code principal en C++, qui est wrappé en module Python avec SWIG. Cette interface web a abouti à la création d'un site de simulations, Simulations@LPMA à l'adresse

http ://simulations.lpma-paris.fr/,

vouée à accueillir les résultats numériques du laboratoire. Cette page se divise en deux branches principales, les applications numériques à vocation didactique et les applications numériques liées aux résultats de recherche des différentes équipes, ou membres du laboratoire.

Le travail d'ingénierie informatique nécessaire à la conception et à la réalisation de ce site ne sera pas détaillé plus avant dans ce manuscrit. Les programmes développés pour l'application "Méthodes Multilevel" de ce site ont néanmoins été massivement mis à contribution dans les diverses sections numériques de cette thèse.

Supposons que $Y_0 \in \mathbf{L}^2(\mathbf{P})$ est une variable aléatoire réelle *non simulable* à coût raisonnable, dont nous voulons calculer l'espérance $I_0 = \mathbf{E}[Y_0]$, une précision $\varepsilon > 0$ étant donnée, en minimisant le coût de simulation. Y_0 n'étant pas simulable, nous supposons avoir à disposition une famille de variables aléatoires réelles $(Y_h)_{h\in\mathcal{H}} \in \mathbf{L}^2(\mathbf{P})$ qui sont elles *simulables* avec un coût petit devant celui de Y_0 et qui approchent Y_0 dans un sens faible et fort comme suit : nous supposons avoir un développement polynomial du biais en une puissance α du paramètre de biais h

$$\mathbf{E}[Y_h] - \mathbf{E}[Y_0] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + o(h^{\alpha \bar{R}}) \tag{1}$$

et nous supposons que l'erreur quadratique est contrôlée par une puissance β de h

$$\left\|Y_h - Y_0\right\|_2^2 = \mathbf{E}\left[\left|Y_h - Y_0\right|^2\right] \leqslant V_1 h^{\beta}. \tag{2}$$

Nous supposons de plus que le coût de simulation de la variable Y_h est inversement linéaire en h, c'est-à-dire $\mathrm{Cost}(Y_h) = \frac{\kappa}{h}$, où κ est une constante qui dépend de nos moyens techniques.

Deux exemples typiques d'application sont les schémas de discrétisation de processus de diffusion et le *nested* Monte Carlo. Dans le premier cas nous désirons calculer $I_0 = \mathbf{E}[Y_0]$, avec $Y_0 = f(X_T)$ et $(X_t)_{t\in[0,T]}$ processus de diffusion. Nous prenons pour approcher Y_0 la variable $Y_h = f(\bar{X}_T^h)$, où $\bar{X}_T^h$ est un schéma de discrétisation de pas $h = \frac{T}{n}$. Sous des hypothèses de régularité de la diffusion de $(X_t)_{t\in[0,T]}$ et de la fonction de payoff f, nous pouvons vérifier que Y_h approche Y_0 au sens fort et faible indiqué par les formules (1) et (2). Dans le deuxième cas nous voulons calculer $I_0 = \mathbf{E}[f(\mathbf{E}[X|Y])]$, $Y_0 = f(\mathbf{E}[X|Y])$, où $X = F(Z,Y)$ avec Z indépendante de Y. On choisit alors pour approcher Y_0 la variable $Y_h = Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^K F(Z_k, Y)\right)$. Sous les bonnes hypothèses sur f et F nous pouvons également vérifier (1) et (2).

Supposons un instant que Y_0 soit simulable, alors nous pourrions prendre $(Y_0^k)_{k\geq 1}$ suite de copies indépendantes de Y_0 et choisir comme estimateur de $I_0 = \mathbf{E}[Y_0]$ un estimateur de Monte Carlo standard

$$I_0^N = \frac{1}{N}\sum_{k=1}^N Y_0^k.$$

Cet estimateur est sans biais : $\mathbf{E}\left[I_0^N\right] - I_0 = 0$, sa variance est donnée par $\mathrm{Var}(I_0^N) = \frac{1}{N}\mathrm{Var}(Y_0)$ et son coût vaut $\kappa_0 N$, où κ_0 est le coût de simulation unitaire de Y_0. Sous la contrainte $\left\|I_0^N - I_0\right\|_2 = \varepsilon$, le choix optimal de la taille N de l'estimateur est donné par

$$N(\varepsilon) = \frac{\mathrm{Var}(Y_0)}{\varepsilon^2},$$

avec un coût optimal qui s'écrit

$$\mathrm{Cost}(I_0^N(\varepsilon)) = \frac{\kappa_0\,\mathrm{Var}(Y_0)}{\varepsilon^2}.$$

Par la suite, il nous sera utile de garder à l'esprit qu'il n'est pas possible d'obtenir mieux que $\mathrm{Cost}(I_\pi^N(\varepsilon)) = C\varepsilon^{-2}$, avec C constante, où nous notons par π l'ensemble des paramètres qui décrivent l'estimateur. En prenant $\pi = h$, l'estimateur Monte Carlo standard (appelé dans la suite Monte Carlo Crude) $I_h^N(\varepsilon)$ se construit à partir de $(Y_h^k)_{k\geq 1}$ suite de copies indépendantes de Y_h et s'écrit

$$I_h^N = \frac{1}{N}\sum_{k=1}^N Y_h^k.$$

Le biais (en utilisant l'hypothèse d'erreur faible (1)), la variance et le coût de cet estimateur s'écrivent

$$\mu(h) = \mathbf{E}[Y_h] - I_0 \simeq c_1 h^\alpha, \quad \mathrm{Var}(I_h^N) = \frac{1}{N}\mathrm{Var}(Y_h), \quad \mathrm{Cost}(I_h^N) = \frac{\kappa}{h}N.$$

Pour atteindre la précision désirée $\left\|I_h^N - I_0\right\|_2^2 = \mu(\varepsilon)^2 + \mathrm{Var}(I_h^N(\varepsilon)) = \varepsilon^2$ nous allons être amenés à choisir $N(\varepsilon) \sim \varepsilon^{-2}$ et $h(\varepsilon) \sim \varepsilon^{\frac{1}{\alpha}}$, ce qui nous conduit à un coût optimal

$$\mathrm{Cost}(I_h^N) \sim \varepsilon^{-2-\frac{1}{\alpha}}.$$

Tous nos efforts seront donc voués à construire à partir de la famille $(Y_h)_{h\in\mathcal{H}}$ des estimateurs I_π^N qui réalisent un coût $C_1\varepsilon^{-2} < \mathrm{Cost}(I_\pi^N) < C_2\varepsilon^{-2-\frac{1}{\alpha}}$, C_1 et C_2 constantes, le plus proche possible du coût sans biais en ε^{-2}.

Nous ne donnons pas ici la définition des estimateurs Multilevel, mais nous anticipons qu'un estimateur Multilevel s'écrit comme la somme de deux termes, un premier terme grossier I_ε^1 qui est un estimateur Monte Carlo Crude qui donne une première approximation de I_0 et un deuxième terme qui contient une somme de couches de corrections de plus en plus fines I_ε^2, qui ajustent l'estimation de I_0 de sorte que l'on obtienne l'erreur ε que l'on s'était fixée.

Dans le premier chapitre de cette thèse, nous donnons les définitions principales des estimateurs Multilevel, avec et sans poids, que nous allons dans la suite noter ML2R et MLMC. Nous allons énoncer les résultats précédents sur les choix de paramètres optimaux qui permettent d'obtenir le coût optimal, qui dépendra de l'ordre d'erreur forte β comme suit

- $\beta > 1$: Pour les deux estimateurs MLMC et ML2R
$$\mathrm{Cost}\left(I_\pi^N(\varepsilon)\right) \precsim K(\alpha,\beta)\varepsilon^{-2}.$$
- $\beta \leq 1$: Le coût diffère pour MLMC et ML2R
$$\mathrm{Cost}\left(I_\pi^N(\varepsilon)\right) \precsim K(\alpha,\beta)\varepsilon^{-2}v(\varepsilon),$$
avec

	$v_{MLMC}(\varepsilon)$	$v_{ML2R}(\varepsilon)$
$\beta = 1$	$\log(1/\varepsilon)^2$	$\log(1/\varepsilon)$
$\beta < 1$	$\varepsilon^{-\frac{1-\beta}{\alpha}}$	$o(\varepsilon^{-\eta})\ \forall\eta > 0$

On obtient donc, pour $\beta \leq 1$, de meilleures performances avec ML2R, l'estimateur Multilevel avec poids. Nous remarquons ici que le calcul des poids dans ML2R est donné par une formule fermée facile à calculer, ce qui suggère d'utiliser ML2R par défaut.

Dans le deuxième chapitre de cette thèse, nous allons analyser le comportement asymptotique des estimateur Multilevel avec et sans poids, en terme de convergence *p.s.* et de convergence en loi. Nous allons énoncer une Loi Forte des Grands Nombres et un Théorème Central Limite.

On remarque que la convergence L^2 vaut par construction, puisque $\left\|I_\pi^N(\varepsilon) - I_0\right\|_2 \leq \varepsilon$. En conséquence, pour toute suite $(\varepsilon_k)_{k\geq 1}$ telle que $\sum_{k\geq 1}\varepsilon_k^2 < +\infty$, il vient

$$\sum_{k\geq 1}\mathbf{E}\left[\left|I_\pi^N(\varepsilon_k) - I_0\right|^2\right] < +\infty,$$

donc $I_\pi^N(\varepsilon_k) \xrightarrow{p.s.} I_0$. Nous pouvons affaiblir l'hypothèse sur la suite $(\varepsilon_k)_{k\geq 1}$ quand Y_h a des moments finis d'ordre plus grand que 2. Plus précisément, sous l'hypothèse d'erreur forte en norme L^p

$$\exists\,\beta > 0, V_1^{(p)} \geqslant 0, \quad \left\|Y_h - Y_0\right\|_p^p = \mathbf{E}\left[\left|Y_h - Y_0\right|^p\right] \leqslant V_1^{(p)}h^{\frac{\beta}{2}p},$$

nous pouvons établir la Loi Forte des Grands Nombres

$$I_\pi^N(\varepsilon_k) \xrightarrow{p.s.} I_0 \quad \text{pour} \quad k \to +\infty,$$

si $\sum_{k\geq 1} \varepsilon_k^p < +\infty$.

Pour le Théorème Central Limite nous allons écrire

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = \frac{\mu(h(\varepsilon), R(\varepsilon))}{\varepsilon} + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 + \frac{\widetilde{I}_\varepsilon^2}{\varepsilon},$$

où $\widetilde{I}_\varepsilon^1$ représente la première couche grossière centrée et $\widetilde{I}_\varepsilon^2$ la somme des couches fines centrées. Le point central de la preuve consiste à montrer

1) $\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_1^2)$,

2) $\dfrac{\widetilde{I}_\varepsilon^2}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_2^2)$.

Nous allons voir que toute la difficulté réside dans la deuxième limite, qui nécessite la vérification d'une condition de Lindeberg. Dans ce but, il sera nécessaire de faire une hypothèse supplémentaire de L^2-uniforme intégrabilité sur les couches de correction.

Nous allons montrer pour conclure ce chapitre sous quelles conditions les deux exemples de référence, les diffusions et le nested, vérifient les hypothèses des deux Théorèmes énoncés.

Ce chapitre a donné lieu à une publication dans la revue Monte Carlo Methods and Applications.

Dans le troisième chapitre nous allons étudier les estimateurs Multilevel dans le cadre des schémas de discrétisation antithétiques. C'est un cadre où nous nous retrouvons avec $\beta > 1$, cadre dans lequel les estimateurs se comportent comme des estimateurs sans biais. Nous nous sommes largement inspirés des résultats de Al Gerbi, Jourdain et Clément dans [AGJC16], dans le soucis d'adapter les estimateur Multilevel avec poids au cadre antithétique.

Le calcul des paramètres optimaux diffère de celui fourni dans [LP17] et repris dans [AGJC16] en raison du fait que l'hypothèse d'erreur forte (2) est remplacée par une hypothèse sur la variance des couches de corrections et que dans le calcul du coût nous devons prendre en considération l'antithéticité des schémas.

Nous allons tester les deux estimateurs Multilevel avec et sans poids sur un modèle de Clark Cameron avec un payoff pour lequel nous avons une formule fermée, ce qui nous permet de vérifier si nous respectons la contrainte sur l'erreur L^2 que nous nous sommes imposés.

Dans le quatrième chapitre nous analysons le *nested* Monte Carlo, en donnant le détail de deux situations, selon la régularité de la fonction de payoff f. Dans les deux cas nous focalisons notre attention sur la preuve des hypothèses d'erreur faible et forte.

Dans le cinquième et dernier chapitre de cette thèse nous nous plaçons à nouveau dans le cadre $\beta > 1$, ce qui nous permet de construire un estimateur sans biais, en randomisant le niveau R d'un estimateur Multilevel. Nous comparons ensuite numériquement cet estimateur aux estimateurs Multilevel avec et sans poids.

Introduction

This thesis has been realized with the support of the Chaire Risques Financiers of the Fondation du Risque of the Institut Louis Bachelier. The aim of this foundation is to bring a sustainable contribution to the development of the french potential in terms of research, teaching and training in all the domain of risk : financial risks, industrial risks, environmental risks, health risks, etc. Since its creation in 2007, the Fondation du Risque gets involved in research and development projects, including the promotion of teaching and research projects and the diffusion of the knowledge through high level training programs. It also promotes the risk teaching through the diffusion of the knowledge for all public, wherther informed or not.

In this manuscript we describe research results that we have obtained during a temporary position of Research and Development Engineer (three years). The development part of this mission focused on the creation of a web interface allowing to compute online Multilevel estimators. This permits in particular an easy and quick illustration of high level research results to a wide, potentially uninformed, public.

The web application is developed in Python, using Flask module, with the addition of HTML and Javascript code. The computation of the parameters and of the estimators themselves is done through the call of a principal C++ code, which is wrapped in a Python module by SWIG. This interface led to the creation of a simulation site, Simulations@LPMA at the address

http ://simulations.lpma-paris.fr/,

conceived to receive various numerical results of the laboratory (and not specifically Multilevel estimators). This page is divided into two main parts, the first one gives several numerical applications in view of teaching purpose and serves as a support for different lectures given at the university. The second part can be seen as a numerical showcase : it exhibits numerical simulations related to research projects and results developed by the different team of the laboratory.

Although the programs developed for the application "Multilevel Methods" of this site have been massively used in the different numerical sections of this thesis, we will not detail in this manuscript the computing engineer work needed for the design and the realization of this web site.

We assume that $Y_0 \in \mathbf{L}^2(\mathbf{P})$ is a random real variable *non simulatable* at a reasonable cost, of which we want to compute an approximation of the expectation $I_0 = \mathbf{E}[Y_0]$, given an accuracy $\varepsilon > 0$, and minimizing the computational cost. Since Y_0 is not simulatable, we assume to have a family of random real variables $(Y_h)_{h\in\mathcal{H}} \in \mathbf{L}^2(\mathbf{P})$ which are *simulatable* with a small cost compared to the cost of Y_0 and that approach Y_0 in a weak and a strong sense as follows : we assume to have a polynomial expansion of the bias in h^α

$$\mathbf{E}[Y_h] - \mathbf{E}[Y_0] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + o(h^{\alpha \bar{R}}) \tag{3}$$

and we assume that the quadratic error is controlled by h^β

$$\left\|Y_h - Y_0\right\|_2^2 = \mathbf{E}\left[\left|Y_h - Y_0\right|^2\right] \leqslant V_1 h^\beta. \tag{4}$$

Moreover, we assume that the simulation cost of the variable Y_h is linear inverse in h, *i.e.* $\text{Cost}(Y_h) = \frac{\kappa}{h}$, where κ is a constant depending on our technical equipment.

Two typical examples of application are the discretization schemes of diffusion processes and the *nested* Monte Carlo. In the first framework we want to compute $I_0 = \mathbf{E}[Y_0]$, with $Y_0 = f(X_T)$ and $(X_t)_{t\in[0,T]}$ diffusion process. We approach Y_0 with the random variable $Y_h = f(\bar{X}_T^h)$, where $\bar{X}_T^h$ is a discretization scheme with step $h = \frac{T}{n}$. Under some regularity assumptions on the diffusion of $(X_t)_{t\in[0,T]}$ and on the payoff function f, we can check that Y_h approaches Y_0 in the weak and strong sense described by the formulas (3) and (4). In the second framework we want to compute $I_0 = \mathbf{E}[f(\mathbf{E}[X|Y])]$, $Y_0 = f(\mathbf{E}[X|Y])$, where $X = F(Z,Y)$ with Z independent of Y. We choose then to approach Y_0 the random variable $Y_h = Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^K F(Z_k, Y)\right)$. Under some good assumption on f and F we can check that we satisfy (3) and (4).

We assume for a moment that Y_0 is simulatable, then we can take a sequence $(Y_0^k)_{k\geq 1}$ of independent copies of Y_0 and choose as an estimator of $I_0 = \mathbf{E}[Y_0]$ a standard Monte Carlo estimator

$$I_0^N = \frac{1}{N}\sum_{k=1}^N Y_0^k.$$

This estimator is unbiased : $\mathbf{E}\left[I_0^N\right] - I_0 = 0$, its variance is given by $\text{Var}(I_0^N) = \frac{1}{N}\text{Var}(Y_0)$ and its cost reads $\kappa_0 N$, where κ_0 is the unitary simulation cost of Y_0. Under the constraint $\left\|I_0^N - I_0\right\|_2 = \varepsilon$, the optimal choice for the size N of the estimator is given by

$$N(\varepsilon) = \frac{\text{Var}(Y_0)}{\varepsilon^2},$$

with an optimal cost

$$\text{Cost}(I_0^N(\varepsilon)) = \frac{\kappa_0\,\text{Var}(Y_0)}{\varepsilon^2}.$$

In what follows, we will keep in mind that it is not possible to get better than the cost $\text{Cost}(I_\pi^N(\varepsilon)) = C\varepsilon^{-2}$, with C constant, where we denote by π the set of parameters that describe the estimator. If we take $\pi = h$, the standard Monte Carlo estimator (called in what follows Crude Monte Carlo) $I_h^N(\varepsilon)$ is built with a sequence $(Y_h^k)_{k\geq 1}$ of independent copies of Y_h and reads

$$I_h^N = \frac{1}{N}\sum_{k=1}^N Y_h^k.$$

The bias (using the weak error assumption (3)), the variance and the cost of this estimator read

$$\mu(h) = \mathbf{E}[Y_h] - I_0 \simeq c_1 h^\alpha, \quad \text{Var}(I_h^N) = \frac{1}{N}\text{Var}(Y_h), \quad \text{Cost}(I_h^N) = \frac{\kappa}{h}N.$$

In order to attain the desired accuracy $\left\|I_h^N - I_0\right\|_2^2 = \mu(\varepsilon)^2 + \text{Var}(I_h^N(\varepsilon)) = \varepsilon^2$ we will need to choose $N(\varepsilon) \sim \varepsilon^{-2}$ and $h(\varepsilon) \sim \varepsilon^{\frac{1}{\alpha}}$, which brings to an optimal cost

$$\text{Cost}(I_h^N) \sim \varepsilon^{-2-\frac{1}{\alpha}}.$$

All of our efforts will be devoted to the construction, starting from the family $(Y_h)_{h\in\mathcal{H}}$, of estimators I_π^N which realize a cost $C_1\varepsilon^{-2} < \text{Cost}(I_\pi^N) < C_2\varepsilon^{-2-\frac{1}{\alpha}}$, C_1 and C_2 constants, as close as possible to ε^{-2}.

We will not give here the definition of the Multilevel estimators, but we anticipate that a Multilevel estimator reads as the sum of two terms, a first coarse term I_ε^1 which is a Crude Monte Carlo giving a first approximation of I_0 and a second term which contains the sum of the correcting levels, finer and finer, I_ε^2, which adjust the estimation of I_0 in order to obtain the error ε that we fixed.

In the first chapter of this thesis, we give the principal definitions of Multilevel estimators, with and without weights, which we note in what follows by ML2R and MLMC. We will recall the previous results, in terms of optimal parameters which lead to the optimal cost, which will depend on the order of the strong error β as follows

- $\beta > 1$: For both estimators MLMC and ML2R

$$\text{Cost}\left(I_\pi^N(\varepsilon)\right) \precsim K(\alpha,\beta)\varepsilon^{-2}.$$

- $\beta \leq 1$: The cost differs for MLMC and ML2R

$$\text{Cost}\left(I_\pi^N(\varepsilon)\right) \precsim K(\alpha,\beta)\varepsilon^{-2}v(\varepsilon),$$

with

	$v_{MLMC}(\varepsilon)$	$v_{ML2R}(\varepsilon)$
$\beta = 1$	$\log(1/\varepsilon)^2$	$\log(1/\varepsilon)$
$\beta < 1$	$\varepsilon^{-\frac{1-\beta}{\alpha}}$	$o(\varepsilon^{-\eta})\ \forall \eta > 0$

Hence we get, for $\beta \leq 1$, better performances with ML2R, the Multilevel estimator with weights. We notice that the computations for the weights in ML2R are given by a closed formula easy to compute, which suggests to use ML2R by default.

In the second chapter of this thesis, we will analyze the asymptotic behaviour of the Multilevel estimators with and without weights, in terms of *a.s.* convergence and convergence in law. We will state a strong law of large numbers and a central limit theorem.

We notice that the L^2 convergence holds by construction , since $\left\|I_\pi^N(\varepsilon) - I_0\right\|_2 \leq \varepsilon$. As a consequence, for all sequence $(\varepsilon_k)_{k\geq 1}$ such that $\sum_{k\geq 1}\varepsilon_k^2 < +\infty$, we get

$$\sum_{k\geq 1}\mathbf{E}\left[\left|I_\pi^N(\varepsilon_k) - I_0\right|^2\right] < +\infty,$$

hence $I_\pi^N(\varepsilon_k) \xrightarrow{p.s.} I_0$. We can weaken the assumption on the sequence $(\varepsilon_k)_{k\geq 1}$ when Y_h has finite moments of order greater than 2. More precisely, under the strong error assumption in L^p norm

$$\exists\, \beta > 0, V_1^{(p)} \geqslant 0, \quad \left\|Y_h - Y_0\right\|_p^p = \mathbf{E}\left[\left|Y_h - Y_0\right|^p\right] \leqslant V_1^{(p)} h^{\frac{\beta}{2}p},$$

we can set a strong law of large numbers

$$I_\pi^N(\varepsilon_k) \xrightarrow{p.s.} I_0 \quad \text{as} \quad k \to +\infty,$$

if $\sum_{k\geq 1}\varepsilon_k^p < +\infty$.

For the central limit theorem we will write

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = \frac{\mu(h(\varepsilon), R(\varepsilon))}{\varepsilon} + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 + \frac{\widetilde{I}_\varepsilon^2}{\varepsilon},$$

where $\widetilde{I}_\varepsilon^1$ represents the first coarse centered and $\widetilde{I}_\varepsilon^2$ the sum of fine centered levels. The central point of the proof consists in showing that

1) $\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_1^2)$,

2) $\dfrac{\widetilde{I}_\varepsilon^2}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_2^2)$.

We will see that the core of the difficulty resides in the second limit, which needs the verification of a Lindeberg condition. For this purpose, it will be necessary to make an additional assumption of L^2-uniform integrability on the correcting levels.

We will conclude by showing under which condition the two reference examples, the diffusions and the nested, satisfy the assumption of these two theorems.

The material of this chapter led to a publication in the journal Monte Carlo Methods and Applications.

In the third chapter we will analyze the Multilevel estimators in the framework of the antithetic discretization schemes. This is a framework where we get $\beta > 1$, hence the estimator behave as unbiased estimators. We have been largely inspired by the results of Al Gerbi, Jourdain et Clément in [AGJC16], in order to adapt the Multilevel estimators with weights to the antithetic framework.

The computation of the optimal parameters differs from the one in [LP17] and recovered in [AGJC16] because the strong error assumption (4) is replaced by an assumption on the variance of the correcting levels and that in the computation of the cost we have to take into account the antitheticity of the schemes.

We will test the two Multilevel estimators with and without weights on a Clark Cameron model with a payoff for which we have a closed formula, in order to check if we respect the constraint on the L^2 error that we fixed.

In the fourth chapter we analyze the *nested* Monte Carlo, giving the detail of two situations, depending on the regularity of the payoff function f. In both cases we focus on the proof of the assumptions of weak and strong error.

In the fifth and last chapter of this thesis we will consider again the setting $\beta > 1$, which allows us to build an unbiased estimator, by randomizing the level of the Multilevel estimators. Then we will compare numerically this estimator to the Multilevel estimators with and without weights.

Chapitre 1

Estimateurs Multilevel

In this chapter we introduce the Multilevel estimators with and without weights and we give a generic form for these estimators, in order to group under a same notation different frameworks, such as standard Multilevel Monte Carlo, introduced by Giles in [Gil08], Multilevel Monte Carlo Richardson-Romberg, introduced by Lemaire and Pagès in [LP17], Multilevel Monte Carlo for diffusions with antithetic discretization schemes, analyzed by Al Gerbi, Jourdain, Clément in [AGJC16] and Randomized Multilevel Monte Carlo, studied by Glynn and Rhee in [RG15]. We retrace the computations that led to the choice of the optimal parameters, giving a generic version of the Theorem of minimization of the cost. This Theorem returns an upper bound of the cost of the type $Kv(\varepsilon)$ with K constant and $v(\varepsilon) \to +\infty$ as $\varepsilon \to 0$, providing a tool of comparison between Multilevel estimators with and without weights, in terms of divergence rate of $v(\varepsilon)$ and giving a sharper multiplicative constant K than in [LP17] (see Theorem 1.3.5 and relative comments).

Dans ce Chapitre nous introduisons les estimateurs Multilevel avec et sans poids et nous donnons une écriture générique pour ces estimateurs, de manière à regrouper sous une même notation différents cadres, tels que Multilevel Monte Carlo standard, introduit par Giles dans [Gil08], Multilevel Monte Carlo Richardson-Romberg, introduit par Lemaire et Pagès dans [LP17], Multilevel Monte Carlo dans le cadre des diffusions avec schémas antithétiques, étudiés par Al Gerbi, Jourdain, Clément dans [AGJC16] et Randomized Multilevel Monte Carlo, étudié par Glynn et Rhee dans [RG15]. Nous allons retracer les calculs qui mènent au choix des paramètres optimaux, en donnant un version générique du Théorème de minimisation du coût. Ce Théorème fourni un majorant du coût des estimateurs multilevel du type $Kv(\varepsilon)$ avec K constante et $v(\varepsilon) \to +\infty$ lorsque $\varepsilon \to 0$, en donnant un outil de comparaison entre les estimateurs Multilevel avec et sans poids, en termes de vitesse de divergence de $v(\varepsilon)$ et de constante K, qui est améliorée par rapport à celle obtenue dans [LP17] (voir Théorème 1.3.5 et commentaires).

1.1 Cadre général

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et $(Y_h)_{h\in\mathcal{H}}$ une famille de variables aléatoires réelles $Y_h \in \mathbf{L}^2(\mathbf{P})$ simulables, avec $\mathcal{H} = \left\{\frac{\mathbf{h}}{n}, n \geq 1\right\}$ et $\mathbf{h} \in \mathbf{R}^+$ un paramètre fixé, approchant une variable aléatoire réelle non simulable $Y_0 \in \mathbf{L}^2(\mathbf{P})$. On veut calculer

$$I_0 = \mathbf{E}\left[Y_0\right],$$

une précision $\varepsilon > 0$ étant donnée et en minimisant le coût de simulation.

Un exemple typique de ce type de situation sont les schémas de discrétisation des processus de diffusion, où l'on veut calculer $\mathbf{E}[Y_0]$ avec $Y_0 = f(X_T)$, où $(X_t)_{t\in[0,T]}$ est un processus de diffusion non simulable que l'on approche par un schéma de discrétisation $Y_h = f(\bar{X}_T^h)$, de pas $h = T/n$.

Un autre exemple, traité par Lemaire et Pagès dans [LP17], est celui de Monte Carlo *nested*, où l'on cherche à calculer $I_0 = \mathbf{E}[f(\mathbf{E}[X|Y])]$, avec $Y_0 = f(\mathbf{E}[X|Y])$. On suppose que $X = F(Z,Y)$ avec Z indépendante de Y et on utilise la famille $Y_h = Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right)$ pour approcher Y_0.

Rappelons que si Y_0 était simulable, nous pourrions construire un estimateur sans biais à partir d'une suite $(Y_0^k)_{k\geq 1}$ de copies indépendantes de Y_0 en écrivant un estimateur de Monte Carlo (appelé dans la suite Crude Monte Carlo)

$$I_0^N = \frac{1}{N}\sum_{k=1}^{N} Y_0^k.$$

Cet estimateur est clairement sans biais $\mathbf{E}\left[I_0^N\right] = I_0$, sa variance vaut $\mathrm{Var}(I_0^N) = \mathrm{Var}(Y_0)/N$ et son coût est $\kappa_0 N$, où κ_0 est une constante qui correspond au temps computationnel de simulation de Y_0 (qui dépend de nos moyens techniques). Sous la contrainte $\left\|I_0^N - I_0\right\|_2 = \varepsilon$, nous obtenons une identification de la taille de l'estimateur $N = \mathrm{Var}(Y_0)/\varepsilon^2$, qui nous ramène à un coût de l'estimateur $\mathrm{Cost}(I_0^N) = \kappa_0 \mathrm{Var}(Y_0)/\varepsilon^2$. Par la suite, nous portons l'attention sur le fait qu'il nous sera impossible d'obtenir un coût inférieur à celui du cadre sans biais.

Nous allons supposer que les variables aléatoires $(Y_h)_{h\in\mathcal{H}}$ approchent Y_0 dans un sens fort et dans un sens faible comme suit :

Hypothèse d'erreur faible (Développement du biais) :

$$\exists\, \alpha > 0, \bar{R} \geqslant 1, (c_r)_{1\leq r\leq \bar{R}}, \quad \mathbf{E}[Y_h] - \mathbf{E}[Y_0] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + o\left(h^{\alpha \bar{R}}\right). \tag{WE_α}$$

Hypothèse d'erreur forte (Erreur quadratique) :

$$\exists\, \beta > 0, V_1 \geqslant 0, \quad \left\|Y_h - Y_0\right\|_2^2 = \mathbf{E}\left[\left|Y_h - Y_0\right|^2\right] \leqslant V_1 h^\beta. \tag{SE_β}$$

De plus, nous supposons que la simulation de Y_h a un coût inversement proportionnel au paramètre de biais, *i.e.* $\mathrm{Cost}(Y_h) = \kappa_Y h^{-1}$, où κ_Y est une constante qui dépend des moyens techniques à disposition.

L'estimateur Crude Monte Carlo, pour un h fixé, s'écrit

$$I_h^N = \frac{1}{N}\sum_{k=1}^{N} Y_h^k, \tag{1.1}$$

où $(Y_h^k)_{k\geq 1}$ sont des copies *i.i.d.* de Y_h et N est la taille de l'estimateur, qui contrôle l'erreur. Nous observons que le biais $\mu(h) = \mathbf{E}\left[I_h^N\right] - I_0$ n'est pas nul et que par l'hypothèse d'erreur faible $\mu(h) \simeq c_1 h^\alpha$. La variance vaut $\mathrm{Var}(I_h^N) = \frac{1}{N}\mathrm{Var}(Y_h)$ et le coût est donné

par $\mathrm{Cost}(I_h^N) = \kappa_Y h^{-1} N$. Une réduction du paramètre de biais h ou une augmentation de la taille N impliquent l'un et l'autre une augmentation du coût de l'estimateur.

L'idée des estimateurs Multilevel, introduits par Giles dans [Gil08], naît des deux remarques suivantes : 1) simuler Y_h est moins coûteux que simuler $Y_{\frac{h}{M}}$ et 2) $\mathbf{E}\left[Y_{\frac{h}{M}}\right] = \mathbf{E}\left[Y_h\right] + \mathbf{E}\left[Y_{\frac{h}{M}} - Y_h\right]$. On construit un estimateur Two-level Monte Carlo comme suit

$$I_\pi^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Y_h^{(1),k} + \frac{1}{N_2}\sum_{k=1}^{N_2}\left(Y_{\frac{h}{M}}^{(2),k} - Y_h^{(2),k}\right)$$

où les suites $\left(Y_h^{(1),k}\right)_{k\geq 1}$ et $\left(Y_h^{(2),k}, Y_{\frac{h}{M}}^{(2),k}\right)_{k\geq 1}$ sont indépendantes et $Y_h^{(2),k}$ et $Y_{\frac{h}{M}}^{(2),k}$ sont construites à partir de la même réalisation aléatoire. Cette version Two-level correspond dans les faits à la méthode de réduction de variance, dite Statistical Romberg method, présentée par Kebaier dans [Keb05], où la couche de correction est utilisée comme une variable de contrôle qui réduit la variance. Le biais est réduit d'un facteur $M^{-\alpha}$ par rapport au biais de l'estimateur Crude Monte Carlo. Nous pouvons imaginer que, la quantité $Y_{\frac{h}{M}}^{(2),k} - Y_h^{(2),k}$ étant petite, sa variance soit petite et qu'en conséquence la taille N_2 de cette couche de correction soit petite. La taille N_1 de la première couche au contraire sera plus grande, mais ceci est compensé par le fait que $Y_h^{(1)}$ est une approximation grossière de Y_0 et donc moins coûteuse. Afin de généraliser cette définition à un niveau quelconque, prenons en considération une profondeur $R \geqslant 2$ (qui sera le niveau le plus fin et donc le plus coûteux de simulation) et une suite géométrique décroissante de paramètres de biais $h_j = h/n_j$ avec $n_j = M^{j-1}$, $j = 1, \ldots, R$. Une fois la taille N de l'estimateur fixée, nous considérons une stratification $q = (q_1, \ldots, q_R)$, telle que pour chaque niveau $j = 1, \ldots, R$, nous allons simuler $N_j = \lceil N q_j \rceil$ scénarios (voir (1.2) et (1.4) ci-dessous). Un estimateur Multilevel est formé par une première couche de simulation grossière de taille N_1, qui correspond à un Monte Carlo classique donnant une première estimation de I_0, suivie d'une somme de couches de correction de plus en plus fines, chacune de taille N_j, $j = 2, \ldots, N_R$, qui corrigent l'erreur faite par la première couche.

1.1.1 Définitions des estimateurs

Considérons R copies indépendantes de la famille $Y^{(j)} = (Y_h^{(j)})_{h\in\mathcal{H}}$, $j = 1, \ldots, R$, rattachées à des copies *indépendantes* $Y_0^{(j)}$ de Y_0. De plus, soient $(Y^{(j),k})_{k\geq 1}$ des suites indépendantes de copies indépendantes de $Y^{(j)}$. Pour un choix de paramètres $\pi = (h, q, R)$ fixé, un estimateur Multilevel Monte Carlo de taille N, introduit par Giles dans [Gil08], s'écrit

$$I_\pi^N = I_{h,R,q}^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R}\frac{1}{N_j}\sum_{k=1}^{N_j}\left(Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k}\right). \tag{1.2}$$

En vertu de l'hypothèse d'erreur faible et de la somme télescopique des couches de corrections, le biais de cet estimateur est réduit d'un facteur $M^{-\alpha(R-1)}$ par rapport au biais de l'estimateur Monte Carlo Crude. Plus précisément

$$\mu(h, R, M) \simeq c_1\left(\frac{h}{M^{R-1}}\right)^\alpha. \tag{1.3}$$

La variance de l'estimateur MLMC vaut

$$\mathrm{Var}(I^N_{h,R,q}) = \frac{1}{N_1}\mathrm{Var}(Y_h) + \sum_{j=2}^{R}\frac{1}{N_j}\mathrm{Var}\left(Y^{(j)}_{h_j} - Y^{(j)}_{h_{j-1}}\right)$$

et son coût, si l'on suppose $\mathrm{Cost}(Y^{(j)}_{h_j}, Y^{(j)}_{h_{j-1}}) = \mathrm{Cost}(Y_{h_j}) + \mathrm{Cost}(Y_{h_{j-1}})$, est donné par

$$\mathrm{Cost}(I^N_{h,R,q}) = \kappa_Y \frac{N}{h}\sum_{j=1}^{R} q_j(n_{j-1}+n_j),$$

où l'on rappele $n_j = M^{j-1}$.

Un estimateur Multilevel avec poids, ou Multilevel Richardson-Romberg (ML2R), se construit en adaptant l'approche Multilevel à l'estimateur multistep Richardson-Romberg, introduit par Pagès dans [Pag07]. L'idée de base est d'utiliser des poids pour tuer les termes successifs du biais dans le développement de l'erreur faible. Grâce à l'hypothèse (WE_α), on écrit, pour tout $\mathbf{w}_i \in \mathbf{R}$, $i = 1, 2, 3$,

$$\begin{aligned}\mathbf{E}\left[\mathbf{w}_1 Y_h + \mathbf{w}_2 Y_{\frac{h}{M}} + \mathbf{w}_3 Y_{\frac{h}{M^2}}\right] &= (\mathbf{w}_1+\mathbf{w}_2+\mathbf{w}_3)\mathbf{E}\left[Y_0\right] \\ &+ \mathbf{w}_1\left(c_1 h^\alpha + c_2 h^{2\alpha} + c_3 h^{3\alpha}\right) \\ &+ \mathbf{w}_2\left(c_1\left(\frac{h}{M}\right)^\alpha + c_2\left(\frac{h}{M}\right)^{2\alpha} + c_3\left(\frac{h}{M}\right)^{3\alpha}\right) \\ &+ \mathbf{w}_3\left(c_1\left(\frac{h}{M^2}\right)^\alpha + c_2\left(\frac{h}{M^2}\right)^{2\alpha} + c_3\left(\frac{h}{M^2}\right)^{3\alpha}\right) + o\left(h^{3\alpha}\right).\end{aligned}$$

Dans le but d'obtenir un estimateur de $I_0 = \mathbf{E}\left[Y_0\right]$ sans termes de biais d'ordre 1 et 2, nous sommes ramenés à résoudre le système de Vandermonde

$$\begin{cases} \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 = 1 \\ \mathbf{w}_1 + \mathbf{w}_2 M^{-\alpha} + \mathbf{w}_3 M^{-2\alpha} = 0 \\ \mathbf{w}_1 + \mathbf{w}_2 M^{-2\alpha} + \mathbf{w}_3 M^{-4\alpha} = 0, \end{cases}$$

ce qui nous donne

$$\mathbf{E}\left[\mathbf{w}_1 Y_h + \mathbf{w}_2 Y_{\frac{h}{M}} + \mathbf{w}_3 Y_{\frac{h}{M^2}}\right] = \mathbf{E}\left[Y_0\right] + \widetilde{\mathbf{w}}_4 c_3 h^{3\alpha} + o(h^{3\alpha}),$$

avec $\widetilde{\mathbf{w}}_4 = \mathbf{w}_1 + \mathbf{w}_2 M^{-3\alpha} + \mathbf{w}_3 M^{-6\alpha}$. Un estimateur Three-step Richardson-Romberg s'écrit alors comme

$$\frac{1}{N}\sum_{k=1}^{N}\left(\mathbf{w}_1 Y_h + \mathbf{w}_2 Y_{\frac{h}{M}} + \mathbf{w}_3 Y_{\frac{h}{M^2}}\right)^{(k)}.$$

Nous passons de l'estimateur Three-step à l'estimateur Three-Level Richardson-Romberg en appliquant une transformation d'Abel, $\mathbf{W}_1 = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 = 1$, $\mathbf{W}_2 = \mathbf{w}_2 + \mathbf{w}_3$ et $\mathbf{W}_3 = \mathbf{w}_3$, telle que

$$\mathbf{W}_1\,\mathbf{E}\left[Y_h\right] + \mathbf{W}_2\,\mathbf{E}\left[Y_{\frac{h}{M}} - Y_h\right] + \mathbf{W}_3\,\mathbf{E}\left[Y_{\frac{h}{M^2}} - Y_{\frac{h}{M}}\right] = \mathbf{E}\left[\mathbf{w}_1 Y_h + \mathbf{w}_2 Y_{\frac{h}{M}} + \mathbf{w}_3 Y_{\frac{h}{M^2}}\right],$$

ce qui donne

$$\frac{\mathbf{W}_1}{N_1}\sum_{k=1}^{N_1} Y_h^{(1),k} + \frac{\mathbf{W}_2}{N_2}\sum_{k=1}^{N_2}\left(Y_{\frac{h}{M}}^{(2),k} - Y_h^{(2),k}\right) + \frac{\mathbf{W}_3}{N_3}\sum_{k=1}^{N_3}\left(Y_{\frac{h}{M^2}}^{(3),k} - Y_{\frac{h}{M}}^{(3),k}\right).$$

En généralisant cette définition à un niveau quelconque, un estimateur Multilevel Richardson-Romberg (ML2R), introduit par Lemaire et Pagès dans [LP17], s'écrit

$$I_\pi^N = I_{h,R,q}^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R}\frac{\mathbf{W}_j^R}{N_j}\sum_{k=1}^{N_j}\left(Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k}\right), \tag{1.4}$$

avec les poids $\left(\mathbf{W}_j^R\right)_{j=1,\dots,R}$ définis comme $\mathbf{W}_j^R = \sum_{k=j}^R \mathbf{w}_k^R$ où les $\mathbf{w}_k^R$, $k=1,\dots,R$ sont les solutions du système de Vandermonde $V\,\mathbf{w}^R = e_1$ avec

$$V = V(M,R,\alpha) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & M^{-\alpha} & \cdots & M^{-\alpha(R-1)} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & M^{-\alpha(R-1)} & \cdots & M^{-\alpha(R-1)^2} \end{pmatrix}. \tag{1.5}$$

Le biais de l'estimateur ML2R est encore une fois réduit par rapport au biais de l'estimateur MLMC et s'écrit

$$\mu(h,R,M) \simeq (-1)^{R-1} c_R \left(\frac{h^R}{M^{\frac{R(R-1)}{2}}}\right)^\alpha. \tag{1.6}$$

Ce gain sur le biais se paye dans le calcul de la variance, dans laquelle apparaissent les poids $\mathrm{Var}(I_{h,R,q}^N) = \frac{1}{N_1}\mathrm{Var}(Y_h) + \sum_{j=2}^{R}\frac{(\mathbf{W}_j^R)^2}{N_j}\mathrm{Var}\left(Y_{h_j}^{(j)} - Y_{h_{j-1}}^{(j)}\right)$. Le coût reste le même que pour MLMC, les poids n'ayant pas d'influence dans le calcul du coût.

1.1.2 Quelques remarques sur les poids

Nous remarquons que les poids sont donnés par une formule fermée et que l'ajout de complexité dans le calcul de l'estimateur est donc négligeable. Nous donnons la formule ci dessous et traitons le cas $c_1 = 0$.

1.1.2.1 Formule fermée

Quand $c_1 \neq 0$, la formule fermée pour les poids, solution du système (1.5), est donnée par

$$\mathbf{w}_k^R = \frac{n_k^{\alpha(R-1)}}{\prod_{j\neq k}(n_k^\alpha - n_j^\alpha)} = \frac{(-1)^{R-k} M^{-\frac{\alpha}{2}(R-k)(R-k+1)}}{\prod_{j=1}^{k-1}(1-M^{-j\alpha})\prod_{j=1}^{R-k}(1-M^{-j\alpha})}. \tag{1.7}$$

1.1.2.2 Cas $c_1 = 0$

Nous remarquons que lorsque le premier coefficient dans le développement du biais est nul, *i.e.* $c_1 = 0$, pour construire un estimateur qui tue les premiers $R-1$ termes dans le développement du biais, de sorte à garder seulement les termes en c_R comme dans la

formule (1.6), nous aurons désormais besoin que de $R-1$ poids, que nous allons noter $\widetilde{\mathbf{w}}_j$, $j=1,\dots,R-1$, au lieu que R comme dans le cas $c_1 \neq 0$.

Nous notons $n_j = M^{j-1}$ et nous remarquons que si $(\mathbf{w}_1,\dots,\mathbf{w}_{R-1})$ est solution du système

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & n_2^{-\alpha} & \cdots & n_{R-1}^{-\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-2)} & \cdots & n_{R-1}^{-\alpha(R-2)} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_{R-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

alors

$$\begin{pmatrix} 1 & n_2^{-\alpha} & \cdots & n_{R-1}^{-\alpha} \\ 1 & n_2^{-2\alpha} & \cdots & n_{R-1}^{-2\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-1)} & \cdots & n_{R-1}^{-\alpha(R-1)} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ n_2^{\alpha}\,\mathbf{w}_2 \\ \vdots \\ n_{R-1}^{\alpha}\,\mathbf{w}_{R-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

ce qui mène à l'écriture des poids suivante, dans le cas $c_1 = 0$,

$$\widetilde{\mathbf{w}}_r^R = \frac{n_r^{\alpha}\,\mathbf{w}_r^{R-1}}{\sum_{j=1}^{R-1} n_j^{\alpha}\,\mathbf{w}_j^{R-1}} \quad r=1,\dots,R-1. \tag{1.8}$$

En remplaçant la valeur des poids $\mathbf{w}$ dans les formules (1.8), le biais de l'estimateur ML2R construit avec les nouveaux $R-1$ poids $\widetilde{\mathbf{w}}$ vaut

$$\mathbf{E}\left[I_\pi^N\right] - I_0 = \frac{1}{\sum_{j=1}^{R-1} n_j^{\alpha}\,\mathbf{w}_j^{R-1}} (-1)^{R-2} c_R \left(\frac{h^R}{M^{\frac{R(R-1)}{2}}}\right)^{\alpha}. \tag{1.9}$$

En valeur absolue, ce biais diffère du biais sans premier coefficient nul (1.6) uniquement par le facteur $\frac{1}{\sum_{j=1}^{R-1} n_j^{\alpha}\,\mathbf{w}_j^{R-1}}$. Analysons la valeur de la quantité $\sum_{j=1}^{R-1} n_j^{\alpha}\,\mathbf{w}_j^{R-1}$. Prenons la fraction rationnelle décomposée en éléments simples

$$\frac{1}{\prod_{j=1}^{R}(x-n_j^{-\alpha})} = \sum_{i=1}^{R} \frac{1}{\prod_{j\neq i}(n_i^{-\alpha}-n_j^{-\alpha})} \frac{1}{(x-n_i^{-\alpha})}.$$

On dérive des deux côtés

$$\sum_{i=1}^{R} \frac{(-1)}{(x-n_i^{-\alpha})^2} \frac{1}{\prod_{j\neq i}(x-n_j^{-\alpha})} = \sum_{i=1}^{R} \frac{1}{\prod_{j\neq i}(n_i^{-\alpha}-n_j^{-\alpha})} \frac{(-1)}{(x-n_i^{-\alpha})^2},$$

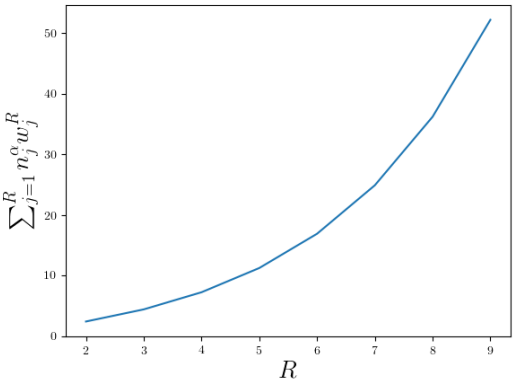
ce qui donne, pour $x=0$,

$$\sum_{i=1}^{R} (-1)^{R+1} n_i^{\alpha}\,\underline{n}!^{\alpha} = \sum_{i=1}^{R} \frac{(-1)^{R-1} n_i^{\alpha(R-1)} n_i^{\alpha}\,\underline{n}!^{\alpha}}{\prod_{j\neq i}(n_j^{\alpha}-n_i^{\alpha})},$$

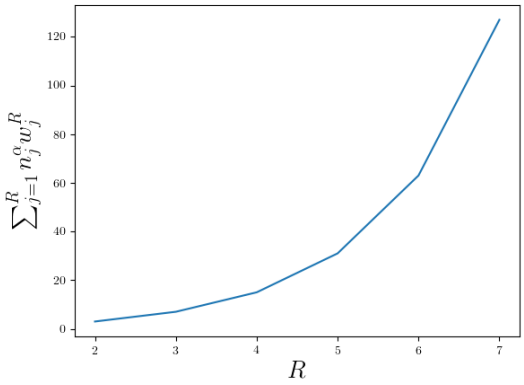
avec $\underline{n}! = \prod_{j=1}^{R} n_j$. En simplifiant et en rappelant la valeur des poids (1.7), nous obtenons

$$\frac{M^{\alpha R}-1}{M^{\alpha}-1} = \sum_{i=1}^{R} n_i^{\alpha} = \sum_{i=1}^{R} n_i^{\alpha}\,\mathbf{w}_i^R,$$

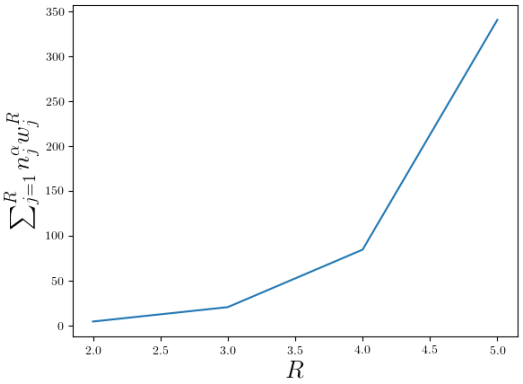
(voir Figures 1.1a, 1.1b et 1.1c). Cette quantité est plus grande que 1 et tend vers l'infini, ce biais est donc inférieur au précédent avec les anciens poids, ce qui implique que, à c_R équivalent, un choix de R qui rendrait le biais (1.6) inférieur à une certaine quantité, rendra aussi le biais (1.9) inférieur à cette même quantité. Cette considération sera à retenir lorsque nous voudrons calculer les paramètres optimaux dans le cas $c_1 = 0$.



(a) $\alpha = 0.5$.



(b) $\alpha = 1$.



(c) $\alpha = 2$.

FIGURE 1.1 – $\sum_{j=1}^{R} n_j^{\alpha} \mathbf{w}_j^R$.

1.2 Nouvelle notation

Nous introduisons une nouvelle notation qui permet de prendre en compte des formes de couches de correction différentes de $Y_{\frac{h}{M}} - Y_h$, comme l'on peut en rencontrer dans les schémas de diffusion antithétiques ou bien dans la version smooth de *nested* (voir les Chapitre 3 et 4). Pour continuer à tirer avantage de la somme télescopique dans le développement du biais, l'estimateur doit rester le même sous le signe d'espérance. Soient $(Y_h^c)_{h\in\mathcal{H}}$ et $\left(Y_h^f\right)_{h\in\mathcal{H}}$ deux familles de variables aléatoires de carré intégrable qui approchent I_0 dans le même sens de l'erreur faible (WE_α), donc

$$\mathbf{E}\left[Y_h^c\right] = \mathbf{E}\left[Y_h^f\right]. \tag{1.10}$$

Nous allons utiliser Y_h^c pour les simulations de pas grossier et Y_h^f pour les pas fins. On suppose que le coût de simulation de chaque famille est inversement proportionnel aux paramètres de biais, plus précisément $\mathrm{Cost}(Y_h^c) = \kappa_c h^{-1}$ et $\mathrm{Cost}(Y_h^f) = \kappa_f h^{-1}$. De plus, on suppose que le coût du couple s'écrit comme une fonction bivariée positivement homogène des coûts marginaux, c'est-à-dire il existe $g : \mathbf{R}^* \times \mathbf{R}^* \to \mathbf{R}^*$ telle que, pour tout $\lambda > 0$, $g(\lambda x, \lambda y) = \lambda g(x, y)$ et $\mathrm{Cost}(Y_{h_{j-1}}^{c,(j)}, Y_{h_j}^{f,(j)}) = g(\mathrm{Cost}(Y_{h_{j-1}}^{c,(j)}), \mathrm{Cost}(Y_{h_j}^{f,(j)}))$. Nous verrons dans la suite que dans le cas des diffusions on a $g(x, y) = x + y$ et dans le cas nested on a $g(x, y) = \max(x, y)$. On note $Y_1^{(c,f)} := Y_{\frac{h}{n_1}}^{(c,f)}$ et on définit

$$Z(h) := Y_{\frac{h}{M}}^f - Y_h^c, \quad Z_j := Z\left(\frac{h}{n_{j-1}}\right),\ j = 2\ldots, R \quad \text{et} \quad Z_1 := Y_1^c.$$

Supposons que les variables Z_j soient indépendantes et que, pour tout $j = 2, \ldots, R$, elles satisfassent l'hypothèse d'erreur forte

$$\exists\, \beta > 0, V_1 \geqslant 0, \quad \left\|Z_j\right\|_2^2 = \mathbf{E}\left[\left|Y_{\frac{h}{n_j}}^f - Y^c_{\frac{h}{n_{j-1}}}\right|^2\right] \leqslant V_1 h^\beta M^{-(j-1)\beta}.$$

Un estimateur MLMC s'écrit comme

$$I_\pi^N = \sum_{j=1}^{R} \frac{1}{N_j} \sum_{k=1}^{N_j} Z_j^k \tag{1.11}$$

et un estimateur ML2R comme

$$I_\pi^N = \sum_{j=1}^{R} \frac{\mathbf{W}_j^R}{N_j} \sum_{k=1}^{N_j} Z_j^k. \tag{1.12}$$

Par la suite nous nous autorisons $N_j = q_j N$. En vertu de l'hypothèse (1.10) faite sur le biais, ces deux estimateurs vérifient les propriétés de réduction du biais (1.3) et (1.6) respectivement pour MLMC et ML2R.

La variance de ces estimateurs est inversement linéaire en N et s'écrit

$$\mathrm{Var}(I_\pi^N) = \frac{1}{N}\nu(\pi) \quad \text{avec} \quad \nu(\pi) = \begin{cases} \displaystyle\sum_{j=1}^{R} \frac{1}{q_j} \mathrm{Var}(Z_j) & \text{pour MLMC,} \\ \displaystyle\sum_{j=1}^{R} \frac{(\mathbf{W}_j^R)^2}{q_j} \mathrm{Var}(Z_j) & \text{pour ML2R.} \end{cases} \tag{1.13}$$

On note $K_j = \mathrm{Cost}(Z_j)$ pour $j \geq 1$ et on écrit

$$K_1 = \kappa_c \frac{n_1}{h} \quad \text{et} \quad K_j = g\left(\kappa_f \frac{n_j}{h}, \kappa_c \frac{n_{j-1}}{h}\right). \tag{1.14}$$

Les coût des estimateurs est linéaire en N et, par homogénéité de la fonction g, s'écrit

$$\mathrm{Cost}(I_\pi^N) = N_1 \kappa_c h^{-1} + \sum_{j=2}^{R} N_j g\left(\kappa_f h^{-1} n_j, \kappa_c h^{-1} n_{j-1}\right) = N\kappa(\pi) \tag{1.15}$$

avec

$$\kappa(\pi) = h^{-1}\left(q_1 \kappa_c + \sum_{j=2}^{R} q_j M^{j-1} g\left(\kappa_f, \kappa_c M^{-1}\right)\right).$$

Une variante de l'estimateur (1.11) permettant de réduire ultérieurement le biais se construit en ajoutant un comportement différent sur la dernière couche (voir [DR15] et [AGJC16]). Nous considérons deux familles supplémentaires de variables aléatoires $(\bar{Y}_h^c)_{h\in\mathcal{H}}$ et $(\bar{Y}_h^f)_{h\in\mathcal{H}}$ telles que $\bar{Y}_h^c$ satisfait la même hypothèse d'erreur faible (WE_α) que Y_h^c et Y_h^f, et $\bar{Y}_h^f$ satisfait une autre hypothèse (WE_α) avec $\bar{\alpha} > \alpha$. On définit

$$\bar{Z}_R := \bar{Y}^f_{\frac{h}{n_R}} - \bar{Y}^c_{\frac{h}{n_{R-1}}}$$

et on suppose que $\bar{Z}_R$ satisfait l'hypothèse d'erreur forte

$$\exists\, \bar{\beta} > 0, \bar{V}_1 \geqslant 0, \quad \left\|\bar{Z}_R\right\|_2^2 = \mathbf{E}\left[\left|\bar{Y}^f_{\frac{h}{n_R}} - \bar{Y}^c_{\frac{h}{n_{R-1}}}\right|^2\right] \leqslant \bar{V}_1 h^{\bar{\beta}} M^{-(j-1)\bar{\beta}}.$$

Un estimateur MLMC s'écrit comme

$$I_\pi^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Z_1^k + \sum_{j=2}^{R-1} \frac{1}{N_j}\sum_{k=1}^{N_j} Z_j^k + \frac{1}{N_R}\sum_{k=1}^{N_R} \bar{Z}_R^k. \tag{1.16}$$

Grâce à la somme télescopique sous le signe d'espérance des termes en Y_h^c, Y_h^f et $\bar{Y}_h^c$, le biais de l'estimateur ne dépend que de la dernière couche et s'écrit

$$\mu(h, R, M) \simeq \bar{c}_1 \left(\frac{h}{M^{R-1}}\right)^{\bar{\alpha}}.$$

La variance et le coût de cet estimateur s'écrivent

$$\mathrm{Var}(I_\pi^N) = \frac{1}{N}\left(\sum_{j=1}^{R-1} \frac{1}{q_j}\mathrm{Var}(Z_j) + \frac{1}{q_R}\mathrm{Var}\left(\bar{Z}_R\right)\right) \tag{1.17}$$

et

$$\mathrm{Cost}(I_\pi^N) = \frac{N}{h}\left(q_1\kappa_c + \sum_{j=2}^{R-1} q_j M^{j-1} g\left(\kappa_f, \kappa_c M^{-1}\right) + q_R M^{R-1} g\left(\bar{\kappa}_f, \bar{\kappa}_c M^{-1}\right)\right). \tag{1.18}$$

L'idée à la base de cette construction est que le gain en termes d'ordre dans développement du biais, $\bar{\alpha} > \alpha$, se paye par une augmentation du coût, $\bar{\kappa}_f > \kappa_f$. En utilisant l'approximation plus coûteuse $\bar{Y}_h^f$ sur toutes les couches, l'impact sur le coût total de l'estimateur

serait conséquent. Puisque cette construction utilise $\bar{Y}_h^f$ seulement sur la dernière couche, elle permet de profiter du meilleur développement de l'erreur faible de $\bar{Y}_h^f$ sans que la pénalisation sur le coût total de l'estimateur soit trop importante.

L'extension de cette variante à ML2R est sans intérêt, puisque le gain sur le biais obtenu grâce à la dernière couche ne dépasse pas le gain en biais obtenu par ML2R. De plus, la construction des poids $\mathbf{W}_j^R$ est étroitement liée au fait que le développement du biais est le même sur toutes le couches, donc pour ML2R nous ne définissons pas de variante en ce sens.

1.3 Paramètres optimaux

Nous rappelons avant tout que l'algorithme de Giles pour le calcul d'un estimateur MLMC qui respecte l'erreur L^2 requise est un algorithme itératif, où l'on part d'un niveau $L = 2$ et on construit d'abord un estimateur MLMC avec tailles des couches données par l'équation

$$N_\ell = \left\lceil 2\varepsilon^{-2}\sqrt{V_\ell/K_\ell}\left(\sum_{\ell=0}^{L}\sqrt{V_\ell K_\ell}\right)\right\rceil, \quad \ell = 0, 1, 2,$$

où V_ℓ est la variance estimée, K_ℓ est le coût de Z_ℓ et la racine est $M = 2$. Ceci assure que la variance estimée de l'estimateur est inférieure à $\frac{1}{2}\varepsilon^2$. On applique ensuite un contrôle sur l'erreur faible visant à assurer que le carré du biais soit inférieur à $\frac{1}{2}\varepsilon^2$ (et donc une erreur L^2 inférieure à ε)

$$\frac{|\mathbf{E}\left[Y_L - Y_{L-1}\right]|}{2^\alpha - 1} < \frac{\varepsilon}{\sqrt{2}}. \tag{1.19}$$

Si (1.19) est respectée on s'arrête, sinon on pose $L = L + 1$ et on continue tant que la condition (1.19) n'est pas atteinte.

L'algorithme obtenu par Lemaire et Pagès dans [LP17] n'est pas itératif, mais consiste à obtenir un jeu de paramètres optimaux qui garantissent a priori la contrainte sur l'erreur L^2. Nous retraçons avec les nouvelles définitions de la Section 1.2 le calcul des ces paramètres optimaux.

Définition 1.3.1. *On définit l'effort d'un estimateur comme le produit de sa variance et de son coût*

$$\phi(I_\pi^N) = \mathrm{Var}(I_\pi^N)\,\mathrm{Cost}(I_\pi^N).$$

L'effort de l'estimateur $\phi(I_\pi^N)$ s'écrit

$$\phi(I_\pi^N) = \left(\sum_{j=1}^{R}\frac{1}{N_j}\mathrm{Var}(Z_j)\right)\left(\sum_{j=1}^{R}N_jK_j\right) = \left(\sum_{j=1}^{R}\frac{1}{q_j}\mathrm{Var}(Z_j)\right)\left(\sum_{j=1}^{R}q_jK_j\right) \tag{1.20}$$

donc l'effort ne dépend pas de N et nous écrivons $\phi(I_\pi^N) = \phi(\pi) = \kappa(\pi)\nu(\pi)$. Le choix des paramètres optimaux (R^*, h^*, q^*, N^*) se fait en trois étapes.

1) *À R fixé.* Choix d'une stratification optimale $q^* = (q_1^*, \ldots, q_R^*)$ qui minimise l'effort.
2) *À R fixé.* Choix du paramètre de biais optimal $h^\star$ qui minimise le coût, sous la contrainte
$$\left\|I_\pi^N - I_0\right\|_2 \leq \varepsilon.$$
La taille optimale N^* découle du choix du biais.

3) Choix de la profondeur optimale $R^*(\varepsilon)$ qui sature la contrainte $h^*(\varepsilon, R^*(\varepsilon)) = \mathbf{h}$.

Le Lemme suivant est d'une utilité cruciale et résout facilement l'étape 1.

Lemme 1.3.2. *Pour tout $j \in \{1, \dots, R\}$, soient $a_j > 0$, $b_j > 0$ tels que $\sum_{j=1}^{R} q_j = 1$. Alors*

$$\left(\sum_{j=1}^{R} \frac{a_j}{q_j}\right)\left(\sum_{j=1}^{R} b_j q_j\right) \geq \left(\sum_{j=1}^{R} \sqrt{a_j b_j}\right)^2$$

et l'égalité est satisfaite si et seulement si $q_j = \mu\sqrt{a_j b_j^{-1}}$, $j = 1, \dots, R$, avec $\mu = \left(\sum_{k=1}^{R} \sqrt{a_k b_k^{-1}}\right)^{-1}$.

Le choix de la stratification optimale $q^* = (q_1^*, \dots, q_R^*)$ est une conséquence du Lemme 1.3.2.

Théorème 1.3.3. *Soit $\pi_0 = (h, R)$. Alors le minimum de l'effort s'écrit*

▷ *MLMC :*

$$\phi(\pi_0, q^*) = \left(\sum_{j=1}^{R} \sqrt{\mathrm{Var}(Z_j) K_j}\right)^2,$$

avec $q_j^*(\pi_0) = \mu^* \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}}$ et μ^* est la constante de normalisation telle que $\sum_{j=1}^{R} q_j^* = 1$.

▷ *ML2R :*

$$\phi(\pi_0, q^*) = \left(\sum_{j=1}^{R} |\,\mathbf{W}_j^R\,| \sqrt{\mathrm{Var}(Z_j) K_j}\right)^2,$$

avec $q_j^*(\pi_0) = \mu^* |\,\mathbf{W}_j^R\,| \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}}$ et μ^* telle que $\sum_{j=1}^{R} q_j^* = 1$.

Démonstration. Il suffit de remplacer $a_j = \mathrm{Var}(Z_j)$ (pour MLMC) ou $a_j = (\mathbf{W}_j^R)^2 \mathrm{Var}(Z_j)$ (pour ML2R) et $b_j = K_j$ dans le Lemme 1.3.2. □

Nous procédons ensuite au choix du paramètre de biais à R fixé, qui résulte de la minimisation du coût en h.

Théorème 1.3.4. *Soit R fixé. Le coût minimal d'un estimateur MLMC avec la stratification q^* est donné par le choix*

$$h^\star = (1 + 2\alpha)^{-\frac{1}{2\alpha}} \left(\frac{\varepsilon}{|c_1|}\right)^{\frac{1}{\alpha}} M^{R-1} \tag{1.21}$$

et s'écrit, lorsque $\varepsilon \to 0$,

$$\inf_{\substack{h \in \mathcal{H} \\ |\mu(h, q^*)| < \varepsilon}} \mathrm{Cost}(Y^N_{\pi_0, q^*}) \sim \left(\frac{(1 + 2\alpha)^{1 + \frac{1}{2\alpha}}}{2\alpha}\right) \frac{|c_1|^{\frac{1}{\alpha}} \mathrm{Var}(Y_0)}{M^{R-1} \varepsilon^{2 + \frac{1}{\alpha}}}.$$

Le coût minimal d'un estimateur ML2R avec la stratification q^* est donné par le choix

$$h^\star = (1+2\alpha R)^{-\frac{1}{2\alpha R}} \left(\frac{\varepsilon}{|c_R|}\right)^{\frac{1}{\alpha R}} M^{\frac{R-1}{2}} \tag{1.22}$$

et s'écrit, lorsque $\varepsilon \to 0$,

$$\inf_{\substack{h\in\mathcal{H}\\ |\mu(h,q^*)|<\varepsilon}} \mathrm{Cost}(Y^N_{\pi_0,q^*}) \sim \left(\frac{(1+2\alpha R)^{1+\frac{1}{2\alpha R}}}{2\alpha R}\right) \frac{|c_R|^{\frac{1}{\alpha R}} \mathrm{Var}(Y_0)}{M^{\frac{R-1}{2}} \varepsilon^{2+\frac{1}{\alpha R}}}.$$

Démonstration. On remarque que le coût s'écrit comme $\mathrm{Cost}(I^N_\pi) = \phi(\pi)/\mathrm{Var}(I^N_\pi)$. La variance à son tour s'écrit $\mathrm{Var}(I^N_\pi) = \mathbf{E}\left[(I^N_\pi - I_0)^2\right] - \mu(h,R,M)^2$. Sous la contrainte $\mathbf{E}\left[(I^N_\pi - I_0)^2\right] = \varepsilon^2$, nous cherchons donc

$$\arg\min_{h\in\mathcal{H}} \frac{\phi(\pi_0,q^*)}{\varepsilon^2 - \mu(h,R,M)^2}.$$

Nous observons grâce à (1.15) et (1.18) que le coût est inversement proportionnel à h. Les hypothèses (WE_α) et (SE_β) impliquent que, lorsque $h \to 0$, $\mathrm{Var}(Y_h) \to \mathrm{Var}(Y_0)$ et $\mathrm{Var}(Z_j) \to 0$. Donc $\lim_{h\to 0} h\phi(h,R,q^*)$ est constante et on cherche

$$\arg\max_{h\in\mathcal{H}} h(\varepsilon^2 - \mu(h,R,M)^2).$$

Comme dans [LP17], nous rappelons que, pour tous réels $a, R' > 0$, la fonction $g_{a,R'}$ définie par $g_{a,R'}(\xi) = \xi(1 - a^2\xi^{2R'})$, $\xi > 0$, satisfait

$$\xi(a,R') := \arg\max_{\xi>0} g_{a,R'}(\xi) = \left((2R'+1)^{\frac{1}{2}} a\right)^{-\frac{1}{R'}} \tag{1.23}$$

et

$$\max_{(0,+\infty)} g_{a,R'} = \frac{2R'}{(2R'+1)^{1+\frac{1}{2R'}}} a^{-\frac{1}{R'}}. \tag{1.24}$$

Nous remarquons aussi pour la suite que

$$\max_{(0,+\infty)} g_{a,R'} \frac{1}{\xi(a,R')} = \frac{2R'}{2R'+1}. \tag{1.25}$$

En utilisant les développements du biais (1.3) et (1.6) et en remplaçant $(R',a) = \left(R\alpha, \frac{|c_R|}{\varepsilon M^{\frac{\alpha}{2}R(R-1)}}\right)$ pour ML2R et $(R',a) = \left(\alpha, \frac{|c_1|}{\varepsilon M^{\alpha(R-1)}}\right)$ pour MLMC dans (1.23) et (1.24), nous obtenons par le calcul les paramètres de biais optimaux (1.21) et (1.22). □

Le choix de la taille de l'estimateur N découle de la linéarité du coût en fonction de N (1.15) et du développement de la variance. En appliquant la contrainte sur l'erreur L^2, on obtient

$$N = \frac{\phi(\pi)}{\kappa(\pi)\,\mathrm{Var}(I^N_\pi)} = \frac{\nu(\pi)}{\varepsilon^2 - \mu(h,R,M)^2},$$

avec $\nu(\pi)$ donné par (1.13). En remplaçant le paramètre de biais optimal $h^\star$ on obtient, grâce à (1.25), pour MLMC

$$N = \left(1 + \frac{1}{2\alpha}\right) \frac{\sum_{j=1}^R \frac{1}{q_j} \mathrm{Var}(Z_j)}{\varepsilon^2}$$

et pour ML2R

$$N = \left(1 + \frac{1}{2\alpha R}\right) \frac{\sum_{j=1}^{R} \frac{(\mathbf{W}_j^R)^2}{q_j} \mathrm{Var}(Z_j)}{\varepsilon^2}.$$

Avec la stratification optimale q^* on obtient, pour MLMC,

$$N^* = \left(1 + \frac{1}{2\alpha}\right) \frac{\sum_{j=1}^{R} \frac{1}{q_j^*} \mathrm{Var}(Z_j)}{\varepsilon^2} = \left(1 + \frac{1}{2\alpha}\right) \frac{\sum_{j=1}^{R} \sqrt{\mathrm{Var}(Z_j) K_j}}{\mu^* \varepsilon^2}$$

et pour ML2R

$$N^* = \left(1 + \frac{1}{2\alpha R}\right) \frac{\sum_{j=1}^{R} \frac{(\mathbf{W}_j^R)^2}{q_j^*} \mathrm{Var}(Z_j)}{\varepsilon^2} = \left(1 + \frac{1}{2\alpha R}\right) \frac{\sum_{j=1}^{R} |\,\mathbf{W}_j^R\,| \sqrt{\mathrm{Var}(Z_j) K_j}}{\mu^* \varepsilon^2}.$$

Rappelons que la constante de normalisation μ^* n'est pas la même pour MLMC et pour ML2R, plus précisément, pour MLMC $\mu^* = \left(\sum_{j=1}^{R} \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}}\right)^{-1}$ et pour ML2R $\mu^* = \left(\sum_{j=1}^{R} |\,\mathbf{W}_j^R\,| \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}}\right)^{-1}$.

De plus, on remarque que pour MLMC

$$N_j = q_j N = \left(1 + \frac{1}{2\alpha}\right) \frac{1}{\varepsilon^2} \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}} \sum_{j=1}^{R} \sqrt{\mathrm{Var}(Z_j) K_j} \tag{1.26}$$

et pour ML2R

$$N_j = q_j N = \left(1 + \frac{1}{2\alpha R}\right) \frac{1}{\varepsilon^2} |\,\mathbf{W}_j^R\,| \sqrt{\frac{\mathrm{Var}(Z_j)}{K_j}} \sum_{j=1}^{R} |\,\mathbf{W}_j^R\,| \sqrt{\mathrm{Var}(Z_j) K_j}.$$

Il reste à donner la formule pour R. Pour l'estimateur ML2R, nous supposons $|c_R|^{\frac{1}{R}} \to \widetilde{c}_\infty$ lorsque $R \to +\infty$ et nous écrivons

$$h(\varepsilon, R) = \left(\frac{1 + 4\alpha}{1 + 2\alpha R}\right)^{\frac{1}{2\alpha R}} \left(\frac{\widetilde{c}_\infty}{|c_R|^{\frac{1}{R}}}\right)^{\frac{1}{\alpha}} e^{\frac{P(R)}{R}} \mathbf{h},$$

avec $P(R) = \frac{R(R-1)}{2} \log(M) - R \log(K) - \frac{1}{\alpha} \log(\sqrt{1 + 4\alpha}/\varepsilon)$, $K = \widetilde{c}_\infty^{\frac{1}{\alpha}}$. Nous remarquons que pour R grand $\left(\frac{1+4\alpha}{1+2\alpha R}\right)^{\frac{1}{2\alpha R}} \left(\frac{\widetilde{c}_\infty}{|c_R|^{\frac{1}{R}}}\right)^{\frac{1}{\alpha}} \sim 1$. Nous notons R_+ la solution positive de $P(R) = 0$, qui sature le biais, et prenons

$$R^* = \lceil R_+ \rceil.$$

En conséquence de ce choix de R^*, nous choisissons comme paramètre de biais optimal la projection de $h(\varepsilon, R^*(\varepsilon))$ sur l'ensemble $\mathcal{H} = \left\{\frac{\mathbf{h}}{n} : n \in \mathbf{N}\right\}$, qui s'écrit $h^*(\varepsilon) = \mathbf{h} \left\lceil \frac{\mathbf{h}}{h(\varepsilon, R^*(\varepsilon))} \right\rceil^{-1}$.

Le même raisonnement s'applique à l'estimateur MLMC.

Le coût avec ce choix de paramètres optimaux est supérieurement borné par une fonction de ε qui dépend de α et β. Plus précisément

$$\mathrm{Cost}\left(I_\pi^N(\varepsilon)\right) \lesssim K(\alpha, \beta, M) v(\varepsilon) \tag{1.27}$$

	$v_{MLMC}(\varepsilon)$	$v_{ML2R}(\varepsilon)$
$\beta = 1$	$\varepsilon^{-2}\log(1/\varepsilon)^2$	$\varepsilon^{-2}\log(1/\varepsilon)$
$\beta < 1$	$\varepsilon^{-2-\frac{1-\beta}{\alpha}}$	$\varepsilon^{-2}e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}}$

TABLE 1.1 – $v(\varepsilon)$ for $\beta \in (0,1]$.

avec $K(\alpha,\beta,M)$ constante et $v(\varepsilon) = \varepsilon^{-2}$ dans le cas $\beta > 1$ et pour $\beta \in (0,1]$, $v(\varepsilon)$ donné par le Tableau 1.1.

Voir [LP17] et le Chapitre 2 pour une description plus complète de l'asymptotique du coût.

Grâce aux hypothèses (WE_α) et (SE_β), pour tout $j \geq 1$, il existe une constante $\widetilde{V}_1$ telle que l'inégalité suivante est vérifiée pour tout $j = 1,\dots,R$

$$\mathrm{Var}(Z_j) \leq \widetilde{V}_1 h_j^\beta,$$

avec

$$\widetilde{V}_1 = \begin{cases} V_1(1+M^{-\frac{\beta}{2}})^2 & \text{si } \beta < 2\alpha, \\ V_1(1+M^{-\frac{\beta}{2}})^2 - c_1^2(1-M^{-\alpha})^2 & \text{si } \beta = 2\alpha. \end{cases}$$

Plus généralement, nous pouvons remplacer l'hypothèse d'erreur forte (SE_β) directement par une hypothèse sur la variance de la couche

$$\mathrm{Var}(Z_j) \leq \bar{V}_1 h_j^\beta, \quad j = 2,\dots,R. \tag{1.28}$$

Dans la pratique, l'estimation des variances $\mathrm{Var}(Z_j)$, $j = 1,\dots,R$ demande un temps de calcul non négligeable. Pour éviter de devoir effectuer un pré-traitement qui serait trop coûteux pour le calcul de la stratification et de la taille de l'estimateur, il est donc préférable de chercher une expression des paramètres optimaux en fonction des $V_j^* = V^* h_j^\beta$ (avec $V^* = \widetilde{V}_1, \bar{V}_1$ selon l'hypothèse qu'on choisit), plutôt qu'en fonction des variances. Si nous avions maximisé les variances par V_j^*, nous aurions obtenu un majorant de l'effort $\bar{\phi}$

$$\phi(\pi_0, q^*) \leq \bar{\phi}(\pi_0, q^*) = \left(\sum_{j=1}^R \sqrt{V_j^* K_j}\right)^2.$$

Nous rappelons la notation

$$f(\varepsilon) \preceq g(\varepsilon) \quad \text{lorsque} \quad \varepsilon \to 0 \quad \Longleftrightarrow \quad \limsup_{\varepsilon\to 0} \frac{f(\varepsilon)}{g(\varepsilon)} = 1.$$

On note (pour plus de détails voir l'Appendix A)

$$a_\ell := \frac{1}{\prod_{1\leq k\leq \ell-1}(1-M^{-k\alpha})}$$

avec la convention $\prod_{k=1}^0(1-M^{-k\alpha}) = 1$, et

$$b_\ell := (-1)^\ell \frac{M^{-\frac{\alpha}{2}\ell(\ell+1)}}{\prod_{1\leq k\leq \ell}(1-M^{-k\alpha})}.$$

Nous pouvons reproduire le même raisonnement pour la minimisation de l'effort et du coût avec $\bar{\phi}$ et finalement obtenir le Théorème suivant.

Théorème 1.3.5. *Supposons*

$$\mathrm{Var}(Z_j) \leq V_j^* = V^* h_j^\beta.$$

(a) Estimateur ML2R : Supposons (WE_α) *et* $\lim_{R\to+\infty} |c_R|^{\frac{1}{R}} = \widetilde{c}_\infty \in (0, +\infty)$. *Un estimateur ML2R vérifie*

$$\mathrm{Cost}(I_\pi^N(\varepsilon)) \preceq K_{ML2R}(\alpha, \beta, M, V^*, \mathbf{h}) v(\beta, \varepsilon)$$

avec

$$v(\varepsilon, \beta) = \begin{cases} \varepsilon^{-2} & si\ \beta > 1, \\ \varepsilon^{-2}(\log(1/\varepsilon)) & si\ \beta = 1, \\ \varepsilon^{-2} e^{\frac{1-\beta}{\alpha}\sqrt{2\log(1/\varepsilon)\log(M)}} & si\ \beta < 1. \end{cases}$$

La constante $K_{ML2R}(\alpha, \beta, M, V^*, \mathbf{h})$ *est donnée par*

$$\begin{cases} \frac{1}{\mathbf{h}} \left(\sqrt{\mathrm{Var}(Y_\mathbf{h})} + \sqrt{V^* g(1, M^{-1})} \frac{M^{\frac{1-\beta}{2}}}{1 - M^{\frac{1-\beta}{2}}} \right)^2 & si\ \beta > 1, \\ V^* g(1, M^{-1}) \frac{2}{\alpha \log(M)} & si\ \beta = 1, \\ V^* g(1, M^{-1}) \left(a_\infty \sum_{j\geq 1} \left| \sum_{\ell=0}^{j-1} b_\ell \right| M^{\frac{\beta-1}{2} j} \right)^2 & si\ \beta < 1. \end{cases}$$

(b) Estimateur MLMC : Supposons (WE_α) *et* $c_1 \neq 0$. *Un estimateur MLMC vérifie*

$$\mathrm{Cost}(I_\pi^N(\varepsilon)) \preceq K_{MLMC}(\alpha, \beta, M, V^*, \mathbf{h}) v(\beta, \varepsilon)$$

avec

$$v(\varepsilon, \beta) = \begin{cases} \varepsilon^{-2} & si\ \beta > 1, \\ \varepsilon^{-2}(\log(1/\varepsilon))^2 & si\ \beta = 1, \\ \varepsilon^{-2-\frac{1-\beta}{\alpha}} & si\ \beta < 1. \end{cases}$$

La constante $K_{MLMC}(\alpha, \beta, M, V^*, \mathbf{h})$ *est donnée par*

$$\begin{cases} \left(1 + \frac{1}{2\alpha}\right) \frac{1}{\mathbf{h}} \left(\sqrt{\mathrm{Var}(Y_\mathbf{h})} + \sqrt{V^* g(1, M^{-1})} \frac{M^{\frac{1-\beta}{2}}}{1 - M^{\frac{1-\beta}{2}}} \right)^2 & si\ \beta > 1, \\ \left(1 + \frac{1}{2\alpha}\right) V^* g(1, M^{-1}) \frac{1}{\alpha^2 \log(M)^2} & si\ \beta = 1, \\ \left(1 + \frac{1}{2\alpha}\right) V^* g(1, M^{-1}) \frac{M^{\beta-1}}{(1 - M^{\frac{\beta-1}{2}})^2} & si\ \beta < 1. \end{cases}$$

La preuve du Théorème suit pas à pas la preuve du Théorème 3.12 dans [LP17], où au lieu que utiliser une majoration des poids $\mathbf{W}_j^R$, nous améliorons la constante $K(\alpha, \beta, M, V^*, \mathbf{h})$ en tirant profit lorsque $R \to +\infty$ d'une sorte de moyennisation des $\left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}$, qui est un résultat obtenu dans le Lemme 2.4.1 du chapitre 2.
Nous remarquons que,

- *Lorsque* $\beta \leq 1$, *le coût de l'estimateur Multilevel avec poids (ML2R) croît a une vitesse inférieure (et donc meilleure) lorsque* $\varepsilon \to 0$ *que le coût de l'estimateur Multilevel Monte Carlo (MLMC).*

- *Lorsque $\beta > 1$, la vitesse des deux estimateurs est la même (et est la même que celle d'un estimateur sans biais), mais la constante $K_{ML2R}(\alpha, \beta, M, V^*, \mathbf{h})$ est plus petite d'un facteur $\left(1 + \frac{1}{2\alpha}\right)$.*

Il est intéressant de remarquer aussi que, lorsque $\beta = 1$, les constantes devant les vitesses sont faciles à comparer et donnent une condition sur M :

$$\left(1 + \frac{1}{2\alpha}\right) V^* g(1, M^{-1}) \frac{1}{\alpha^2 \log(M)^2} \leq V^* g(1, M^{-1}) \frac{2}{\alpha \log(M)}$$

si et seulement si

$$M \geq e^{\frac{2\alpha+1}{4\alpha^2}}.$$

Quand $\alpha = 1$, par exemple, on obtiens que lorsque $M \geq e^{\frac{3}{4}} \sim 2.117$,

$$K_{MLMC}(\alpha, \beta, M, V^*, \mathbf{h}) \leq K_{ML2R}(\alpha, \beta, M, V^*, \mathbf{h}).$$

Les Paramètres optimaux sont donnés explicitement dans le Tableaux (1.2) et (1.3).

$R(\varepsilon)$	$\left\lceil \frac{1}{2} + \frac{\log\left(\tilde{c}_\infty^{\frac{1}{\alpha}} \mathbf{h}\right)}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log\left(\tilde{c}_\infty^{\frac{1}{\alpha}} \mathbf{h}\right)}{\log(M)}\right)^2 + 2\frac{\log\left(\sqrt{1+4\alpha}/\varepsilon\right)}{\alpha \log(M)}} \right\rceil$
$h(\varepsilon)$	$\mathbf{h} / \left\lceil \mathbf{h}(1+2\alpha R)^{\frac{1}{2\alpha R}} \tilde{c}_\infty^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil$
$q(\varepsilon)$	$q_j = \mu^* \lvert \mathbf{W}_j^R \rvert \sqrt{\frac{V_j^*}{K_j}}\ j = 1, \ldots, R; \sum_{1 \leqslant j \leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha R}\right) \frac{1}{\mu^* \varepsilon^2} \sum_{j=1}^{R} \lvert \mathbf{W}_j^R \rvert \sqrt{V_j^* K_j}$

TABLE 1.2 – Paramètres optimaux pour estimateur ML2R.

$R(\varepsilon)$	$\left\lceil 1 + \frac{\log\left(\lvert c_1 \rvert^{\frac{1}{\alpha}} \mathbf{h}\right)}{\log(M)} + \frac{\log(\sqrt{1+2\alpha}/\varepsilon)}{\alpha \log(M)} \right\rceil$
$h(\varepsilon)$	$\mathbf{h} / \left\lceil \mathbf{h}(1+2\alpha)^{\frac{1}{2\alpha}} \lvert c_1 \rvert^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha}} M^{-(R-1)} \right\rceil$
$q(\varepsilon)$	$q_j = \mu^* \sqrt{\frac{V_j^*}{K_j}}\ j = 1, \ldots, R; \sum_{1 \leqslant j \leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha}\right) \frac{1}{\mu^* \varepsilon^2} \sum_{j=1}^{R} \sqrt{V_j^* K_j}$

TABLE 1.3 – Paramètres optimaux pour estimateur MLMC.

Chapitre 2

Théorèmes limite

We will analyze the asymptotic behaviour of the estimators $\hat{I}_\varepsilon$ of $I_0 = \mathbf{E}(Y_0)$ of Multilevel type, when the required accuracy goes to 0. This is in the optimized framework (where the level of the estimator depends on ε) which ensures the asymptotic minimization of the complexity.

These estimators satisfy a strong law of large numbers, in the sense that for all sequence $(\varepsilon_k)_{k\geq 1}$ such that $\sum_{k\geq 1} \varepsilon_k^2 < +\infty$,

$$I_\pi^N(\varepsilon_k) \xrightarrow{a.s.} I_0, \quad \text{as } k \to +\infty.$$

We will weaken the assumptions on $(\varepsilon_k)_{k\geq 1}$ when Y_h has finite moments of order greater than 2.

Then we are going to show a central limit theorem, using the following decomposition

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = m(\varepsilon) + \sigma_2 \zeta_2^\varepsilon + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}} \sigma_1 \zeta_1^\varepsilon, \quad \text{as } \varepsilon \to 0,$$

where ζ_1^ε and ζ_2^ε are two independent variables such that $(\zeta_1^\varepsilon, \zeta_2^\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_2)$ as $\varepsilon \to 0$ and $m(\varepsilon) = \frac{\mu(\varepsilon)}{\varepsilon}$ with $\mu(\varepsilon) = \mathbf{E}\left[I_\pi^N\right] - I_0$ bias of the estimator. The term $\sigma_2\zeta_2^\varepsilon$ comes from the sum of the fine correcting levels, whereas the term $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sigma_1\zeta_1^\varepsilon$ comes from the first coarse level. We will distinguish between two cases, depending on the value of the limit $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}$, which depends on β. When $\beta > 1$, $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}$ converge to a non null constant, hence both the terms contribute to the asymptotic variance of the estimator. Whereas when $\beta \leq 1$, $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}} \to 0$ and only the variance of the correcting levels contributes to the asymptotic variance of the estimator. The normalized ML2R estimator is asymptotically without bias, *i.e.* $m(\varepsilon) \to 0$, whereas the asymptotic bias of the normalized MLMC is lower and upper bounded $\frac{M^{-\alpha}}{\sqrt{1+2\alpha}} < \lim_{\varepsilon\to 0} m(\varepsilon) \leq \frac{1}{\sqrt{1+2\alpha}}$.

We will display two frameworks where these theorems apply, the discretization schemes for diffusion processes, in the spirit of the results obtained by Ben Alaya and Kebaier in [BAK15], and the nested Monte Carlo, which is a new result to our knowledge.

This chapter led to an article accepted for publication in the journal Monte Carlo Methods and Applications, entitled "Limit theorems for weighted and regular Multilevel estimators" and written with Vincent Lemaire and Gilles Pagès.

Il s'agit d'étudier le comportement asymptotique pour les estimateurs $\hat{I}_\varepsilon$ de $I_0 = \mathbf{E}(Y_0)$ de type Multilevel, lorsque la précision requise tend vers 0, ce dans le cadre optimisé (où, notamment, la profondeur de l'estimateur dépend elle-même de ε) qui assure la minimisation asymptotique de la complexité.

Ces estimateurs obéissent à une loi forte des grands nombres, au sens où pour toute suite $(\varepsilon_k)_{k\geq 1}$ telle que $\sum_{k\geq 1} \varepsilon_k^2 < +\infty$,

$$I_\pi^N(\varepsilon_k) \xrightarrow{p.s.} I_0, \quad \text{lorsque } k \to +\infty.$$

Nous allons pouvoir affaiblir l'hypothèse sur la suite $(\varepsilon_k)_{k\geq 1}$ lorsque Y_h a des moments finis d'ordre supérieur à 2.

Nous allons ensuite montrer un Théorème Central Limite, en utilisant la décomposition suivante

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = m(\varepsilon) + \sigma_2 \zeta_2^\varepsilon + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}} \sigma_1 \zeta_1^\varepsilon, \quad \text{lorsque } \varepsilon \to 0,$$

où ζ_1^ε et ζ_2^ε sont deux variables indépendantes telles que $(\zeta_1^\varepsilon, \zeta_2^\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_2)$ lorsque $\varepsilon \to 0$ et $m(\varepsilon) = \frac{\mu(\varepsilon)}{\varepsilon}$ avec $\mu(\varepsilon) = \mathbf{E}\left[I_\pi^N\right] - I_0$ biais de l'estimateur. Le terme $\sigma_2\zeta_2^\varepsilon$ vient de la somme des couches fines correctives, tandis que le terme $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sigma_1\zeta_1^\varepsilon$ vient de la première couche grossière. Nous allons distinguer deux cas, selon la valeur de la limite de $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}$, qui dépend de β. Lorsque $\beta > 1$, $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}}$ converge vers une constante non nulle, donc les deux couches contribuent à la variance asymptotique de l'estimateur. Tandis que lorsque $\beta \leq 1$, $\frac{1}{\varepsilon\sqrt{N(\varepsilon)}} \to 0$ et seulement la variance des couches correctives contribue à la variance asymptotique de l'estimateur. L'estimateur ML2R est asympotiquement sans biais, c'est-à-dire $m(\varepsilon) \to 0$, tandis que le biais asymptotique de l'estimateur MLMC est encadré $\frac{M^{-\alpha}}{\sqrt{1+2\alpha}} < \lim_{\varepsilon\to 0} m(\varepsilon) \leq \frac{1}{\sqrt{1+2\alpha}}$.

Nous allons exhiber deux contextes où ces Théorèmes s'appliquent, les schémas de discrétisation de processus de diffusion, dans l'esprit des résultats obtenus par Ben Alaya et Kebaier dans [BAK15], et le nested Monte Carlo, résultat nouveau à notre connaissance.

Ce Chapitre a fait l'objet d'un article accepté pour publication dans la revue Monte Carlo Methods and Applications, intitulé "Limit theorems for weighted and regular Multilevel estimators" et co-écrit avec Vincent Lemaire et Gilles Pagès.

Abstract

We aim at analyzing in terms of $a.s.$ convergence and weak rate the performances of the Multilevel Monte Carlo estimator (MLMC) introduced in [Gil08] and of its weighted version, the Multilevel Richardson Romberg estimator (ML2R), introduced in [LP17]. These two estimators permit to compute a very accurate approximation of $I_0 = \mathbf{E}[Y_0]$ by a Monte Carlo type estimator when the (non-degenerate) random variable $Y_0 \in L^2(\mathbf{P})$ cannot be simulated (exactly) at a reasonable computational cost whereas a family of simulatable approximations $(Y_h)_{h\in\mathcal{H}}$ is available. We will carry out these investigations in an abstract framework before applying our results, mainly a Strong Law of Large Numbers and a Central Limit Theorem, to some typical fields of applications : discretization schemes of diffusions and nested Monte Carlo.

2.1 Introduction

In recent years, there has been an increasing interest in Multilevel Monte Carlo approach which delivers remarkable improvements in computational complexity in comparison with standard Monte Carlo in biased framework. We refer the reader to [Gil15] for a broad outline of the ideas behind the Multilevel Monte Carlo method and various recent generalizations and extensions. In this paper we establish a Strong Law of Large Numbers and Central Limit Theorem for two kinds of multilevel estimators, Multilevel Monte Carlo estimator (MLMC) introduced by Giles in [Gil08] and the Multilevel Richardson-Romberg (weighted) estimator introduced in [LP17]. We consider a rather general and in some way abstract framework which will allow us to state these results whatever the strong rate parameter is (usually denoted by β). To be more precise we will deal with the versions of these estimators designed to achieve a root mean squared error (RMSE) ε and establish these results as $\varepsilon \to 0$. Doing so we will retrieve some recent results established in [BAK15] in the framework of Euler discretization schemes of Brownian diffusions. We will also deduce a SLLN and a CLT for Multilevel nested Monte Carlo, which are new results to our knowledge. More generally our result applies to any implementation of Multilevel Monte Carlo methods.

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space and let $(Y_h)_{h\in\mathcal{H}}$ be a family of real-valued random variables in $L^2(\mathbf{P})$ associated to Y_0 where $\mathcal{H} = \left\{\frac{\mathbf{h}}{n}, n \geq 1\right\}$ such that $\lim_{h\to 0} \left\|Y_h - Y_0\right\|_2 = 0$. In the sequel, a fixed $h \in \mathcal{H}$ will be called *bias parameter* (though it appears in a different framework as a discretization parameter). In what follows we will be interested in the computational cost of the estimators denoted by the $\mathrm{Cost}(\cdot)$ function. We assume that the simulation of Y_h has an inverse linear complexity *i.e.* $\mathrm{Cost}(Y_h) = h^{-1}$. A natural estimator of $I_0 = \mathbf{E}[Y_0]$ is the standard Monte Carlo estimator, which reads for a fixed h

$$I_h^N = \frac{1}{N}\sum_{k=1}^{N} Y_h^k \quad \text{with} \quad \mathrm{Cost}(I_h^N) = h^{-1}N, \tag{2.1}$$

where $(Y_h^k)_{k\geq 1}$ are *i.i.d.* copies of Y_h and N is the size of the estimator, which controls the statistical error. In order to give the definition of a Multilevel estimator, we consider a *depth* $R \geqslant 2$ (the finest level of simulation) and a geometric decreasing sequence of bias parameters $h_j = h/n_j$ with $n_j = M^{j-1}$, $j = 1, \ldots, R$. If N is the estimator size, we consider

an allocation policy $q = (q_1, \dots, q_R)$, such that, at each level $j = 1, \dots, R$, we will simulate $N_j = \lceil N q_j \rceil$ scenarios (see (2.2) and (2.3) below). Thus, we consider R independent copies of the family $Y^{(j)} = (Y_h^{(j)})_{h \in \mathcal{H}}$, $j = 1, \dots, R$, attached to *independent* random copies $Y_0^{(j)}$ of Y_0. Moreover, let $(Y^{(j),k})_{k \geq 1}$ be independent sequences of independent copies of $Y^{(j)}$. We denote by I_π^N an estimator of size N of I_0, attached to a simulation parameter $\pi \in \Pi \subset \mathbf{R}^d$.

▷ A standard Multilevel Monte Carlo (MLMC) estimator, as introduced by Giles in [Gil08], reads

$$I_\pi^N = I_{h,R,q}^N = \frac{1}{N_1} \sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R} \frac{1}{N_j} \sum_{k=1}^{N_j} \left(Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k} \right) \tag{2.2}$$

with $\pi = (h, R, q)$.

▷ A Multilevel Richardson Romberg (ML2R) estimator, as introduced in [LP17], is a weighted version of (2.2) which reads

$$I_\pi^N = I_{h,R,q}^N = \frac{1}{N_1} \sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R} \frac{\mathbf{W}_j^R}{N_j} \sum_{k=1}^{N_j} \left(Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k} \right) \tag{2.3}$$

with $\pi = (h, R, q)$. The weights $\left(\mathbf{W}_j^R\right)_{j=1,\dots,R}$ are explicitly defined as functions of the weak error rate α (see equation ($WE_{\alpha,\bar{R}}$) below) and of the refiners n_j, $j = 0, \dots, R$ in order to kill the successive bias terms in the weak error expansion (see Section 2.4.3 for more details on the weights). When no ambiguity, we will keep denoting by I_π^N estimators for both classes. We notice that a Crude Monte Carlo estimator of size N formally appears as an ML2R estimator with $R = 1$ and an MLMC estimator appears as an ML2R estimator in which the weights set $\mathbf{W}_j^R = 1$, $j = 1, \dots, R$. Based on the inverse linear complexity of Y_h, it is clear that the simulation cost of both MLMC and ML2R estimators is given by

$$\mathrm{Cost}(I_{h,R,q}^N) = \frac{N}{h} \sum_{j=1}^{R} q_j (n_{j-1} + n_j)$$

with the convention $n_0 = 0$. The difference between the cost of MLMC and of ML2R estimator comes from the different choice of the parameters M, R, h, q and N.

The calibration of the parameters is the result, a root $M \geqslant 2$ being fixed, of the minimization of the simulation cost, for a given target Mean Square Error or L^2-error ε, namely,

$$(\pi(\varepsilon), N(\varepsilon)) = \underset{\left\| I_\pi^N - I_0 \right\|_2 \leq \varepsilon}{\arg\min} \ \mathrm{Cost}(I_\pi^N). \tag{2.4}$$

This calibration has been done in [LP17] for both estimators MLMC and ML2R under the following assumptions on the sequence $(Y_h)_{h \in \mathcal{H}}$. The first one, called *bias error expansion* (or *weak error assumption*), states

$$\exists\, \alpha > 0, \bar{R} \geqslant 1, (c_r)_{1 \leq r \leq \bar{R}}, \quad \mathbf{E}\left[Y_h\right] - \mathbf{E}\left[Y_0\right] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + h^{\alpha \bar{R}} \eta_{\bar{R}}(h), \tag{$WE_{\alpha,\bar{R}}$}$$

with $\lim_{h \to 0} \eta_{\bar{R}}(h) = 0$. The second one, called *strong approximation error assumption*, states

$$\exists\, \beta > 0, V_1 \geqslant 0, \quad \left\| Y_h - Y_0 \right\|_2^2 = \mathbf{E}\left[\left| Y_h - Y_0 \right|^2 \right] \leqslant V_1 h^\beta. \tag{SE_β}$$

Note that the strong error assumption can be sometimes replaced by the sharper

$$\exists\, \beta > 0, V_1 \geqslant 0, \quad \left\|Y_h - Y_{h'}\right\|_2^2 = \mathbf{E}\left[\left|Y_h - Y_{h'}\right|^2\right] \leqslant V_1 |h - h'|^\beta, \quad h, h' \in \mathcal{H}\,.$$

From now on, we set $I_\pi^N(\varepsilon) := I_{\pi(\varepsilon)}^{N(\varepsilon)}$, where $\pi(\varepsilon)$ and $N(\varepsilon)$ are close to solutions of (2.4) (see [LP17] for the construction of these parameters and Tables 2.1 and 2.2 for the explicit values). As mentioned by Duffie and Glynn in [DG95], the global cost of the standard Monte Carlo with these *optimal* parameters satisfies

$$\mathrm{Cost}\left(I_\pi^N(\varepsilon)\right) \lesssim K(\alpha)\varepsilon^{-(2+\frac{1}{\alpha})}$$

where the finite real constant $K(\alpha)$ depends on the structural parameters $\alpha, \mathrm{Var}(Y_0), \mathbf{h}$ and we recall that $f(\varepsilon) \lesssim g(\varepsilon)$ if and only if $\limsup_{\varepsilon\to 0} g(\varepsilon)/f(\varepsilon) \leq 1$. Giles for MLMC in [Gil08] and Lemaire and Pagès for ML2R in [LP17] showed that, using these *optimal* parameters the global cost is upper bounded by a function of ε, depending on the weak error expansion rate α and on the strong error rate β. More precisely, for both estimators we have

$$\mathrm{Cost}\left(I_\pi^N(\varepsilon)\right) \lesssim K(\alpha, \beta, M) v(\varepsilon) \tag{2.5}$$

where the finite real constant $K(\alpha, \beta, M)$ is explicit and differs between MLMC and ML2R (see [LP17] for more details). Denoting v_{MLMC} and v_{ML2R} the dominated function in (2.5) for the MLMC and ML2R estimator respectively, we obtain two distinct cases. In the case $\beta > 1$ both estimators behaves very well as an unbiased Monte Carlo estimator *i.e.* $v_{MLMC}(\varepsilon) = v_{ML2R}(\varepsilon) = \varepsilon^{-2}$. In the case $\beta \leqslant 1$, the ML2R is asymptotically quite better than MLMC since $\lim_{\varepsilon\to 0} \frac{v_{ML2R}}{v_{MLMC}} = 0$. More precisely, we have the following scheme

	$v_{MLMC}(\varepsilon)$	$v_{ML2R}(\varepsilon)$
$\beta = 1$	$\varepsilon^{-2}\log(1/\varepsilon)^2$	$\varepsilon^{-2}\log(1/\varepsilon)$
$\beta < 1$	$\varepsilon^{-2-\frac{1-\beta}{\alpha}}$	$\varepsilon^{-2} e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}}$

The aim of this paper is to prove a Strong Law of Large Numbers (SLLN) and a Central Limit Theorem (CLT) for both estimators MLMC and ML2R calibrated using these *optimal* parameters. First notice that as these parameters have been computed under the constraint $\left\|I_\pi^N(\varepsilon) - I_0\right\|_2 \leq \varepsilon$, the convergence in L^2 holds by construction. As a consequence, it is straightforward that, for every sequence $(\varepsilon_k)_{k\geq 1}$ such that $\sum_{k\geq 1}\varepsilon_k^2 < +\infty$,

$$\sum_{k\geq 1} \mathbf{E}\left[\left|I_\pi^N(\varepsilon_k) - I_0\right|^2\right] < +\infty, \tag{2.6}$$

so that

$$I_\pi^N(\varepsilon_k) \xrightarrow{a.s.} I_0, \quad \text{as } k \to +\infty.$$

We will weaken the assumption on the sequence $(\varepsilon_k)_{k\geq 1}$ when Y_h has higher finite moments, so we will investigate some L^p criterions for $p \geq 2$. Moreover, provided a sharper strong error assumption and adding some more hypothesis of uniform integrability, we will show that

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} - m(\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right), \quad \text{as } \varepsilon \to 0,$$

with $m(\varepsilon) = \frac{\mu(\varepsilon)}{\varepsilon}$ where $\mu(\varepsilon) = \mathbf{E}\left[I_\pi^N\right] - I_0$ is the bias of the estimator, and $m^2 + \sigma^2 \leq 1$, owing to the explicit expression of the constraint

$$\left\|I_\pi^N(\varepsilon) - I_0\right\|_2^2 = \mu(\varepsilon)^2 + \mathrm{Var}(I_\pi^N(\varepsilon)) \leq \varepsilon^2. \tag{2.7}$$

In particular we will prove that $\lim_{\varepsilon\to 0} m(\varepsilon) = 0$ for the ML2R estimator. More precisely we will use in the proof the expansion

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = m(\varepsilon) + \sigma_2\zeta_2^\varepsilon + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sigma_1\zeta_1^\varepsilon, \quad \text{as } \varepsilon \to 0,$$

where ζ_1^ε and ζ_2^ε are two independent variables such that $(\zeta_1^\varepsilon, \zeta_2^\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_2)$ as $\varepsilon \to 0$. We will see that ζ_1^ε comes from the coarse level of the estimator, while ζ_2^ε derives from the sum of the refined levels. When $\beta > 1$, $\varepsilon\sqrt{N(\varepsilon)}$ converges to a constant, hence the variance σ^2 results from the sum of the variance of the first coarse level σ_1^2 and the variance of the sum of the refined fine levels σ_2^2. When $\beta \in (0, 1]$, since $\varepsilon\sqrt{N(\varepsilon)}$ diverges, the contribution to σ^2 of the coarse level disappears and only the variance of the refined levels contributes to σ^2. More details on m and σ will follow in Section 2.3.

The paper is organized as follows. In Section 2.2 we briefly recall the technical background for Multilevel Monte Carlo estimators. In Section 2.3 we state our main results : a Strong Law of Large Numbers and a Central Limit Theorem in a quite general framework. Section 2.4 is devoted to the analysis of the asymptotic behaviour of the optimal parameters, to the study of the weights of the ML2R estimator and to the bias of the estimators and its robustness. These are auxiliary results that we need for the proof of the main theorems, which we detail in Section 2.5. In Section 2.6 we apply these results first to the discretization schemes of Brownian diffusions, where we retrieve recent results by Ben Alaya and Kebaier in [BAK15], and secondly to Nested Monte Carlo.

Notations :
• Let $\mathbf{N}^* = \{1, 2, \ldots\}$ denote the set of positive integers and $\mathbf{N} = \mathbf{N}^* \cup \{0\}$.
• For every $x \in \mathbf{R}_+ = [0, +\infty)$, $\lceil x \rceil$ denotes the unique $n \in \mathbf{N}^*$ satisfying $n - 1 < x \leqslant n$.
• If $(a_n)_{n\in\mathbf{N}}$ and $(b_n)_{n\in\mathbf{N}}$ are two sequences of real numbers, $a_n \sim b_n$ if $a_n = \varepsilon_n b_n$ with $\lim_n \varepsilon_n = 1$, $a_n = O(b_n)$ if $(\varepsilon_n)_{n\in\mathbf{N}}$ is bounded and $a_n = o(b_n)$ if $\lim_n \varepsilon_n = 0$.
We denote by • $\mathrm{Var}(X)$ and $\sigma(X)$ the variance and the standard deviation of a random variable X respectively.

2.2 Brief background on MLMC and ML2R estimators

We follow [LP17] and recall briefly the construction of the *optimal* parameters derived from the optimization problem (2.4). The first step is a stratification procedure allowing us to establish the optimal allocation policy $(q_1, \ldots, q_R)$ when the other parameters R, h, M are fixed. We focus now on the *effort* of the estimator defined as the product of the cost times the variance *i.e.* $\mathrm{Effort}(I_\pi^N) = \mathrm{Cost}(I_\pi^N)\,\mathrm{Var}(I_\pi^N)$. Introducing the notations

$$\forall j \geqslant 2,\ Z_j = \left(\frac{h}{M^{j-1}}\right)^{-\frac{\beta}{2}} \left(Y^{(j)}_{\frac{h}{M^{j-1}}} - Y^{(j)}_{\frac{h}{M^{j-2}}}\right) \quad \text{and} \quad Z_1 = Y_h^{(1)},$$

a Multilevel estimator MLMC (2.2) or ML2R (2.3) writes

$$I_\pi^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Z_1^k + \sum_{j=2}^{R}\frac{1}{N_j}\sum_{k=1}^{N_j} W_j^R \left(\frac{h}{M^{j-1}}\right)^{\frac{\beta}{2}} Z_j^k$$

where $W_j^R = 1$ for the MLMC and $W_j^R = \mathbf{W}_j^R$ for the ML2R. By definition and using the approximation $N_j \simeq N q_j$ the effort satisfies

$$\text{Effort}(I_\pi^N) \simeq \left(\sum_{j=1}^{R} q_j \,\text{Cost}(Z_j)\right)\left(\frac{\text{Var}(Y_h)}{q_1} + \sum_{j=2}^{R}(W_j^R)^2\left(\frac{h}{M^{j-1}}\right)^{\beta}\frac{\text{Var}(Z_j)}{q_j}\right). \tag{2.8}$$

Given R, h, M, a minimization of $q \in (0,1)^R \mapsto \text{Effort}(I_\pi^N)$ on $\left\{q \in (0,1)^R, \sum_{j=1}^{R} q_j = 1\right\}$ gives the solution

$$\begin{cases} q_1^* = \mu^* \sqrt{\frac{\text{Var}(Y_h)}{\text{Cost}(Y_h)}} \\ q_j^* = \mu^* \left(\frac{h}{M^{j-1}}\right)^{\frac{\beta}{2}} |W_j^R| \sqrt{\frac{\text{Var}(Z_j)}{\text{Cost}(Z_j)}} \end{cases} \quad \text{with } \mu^* \text{ such that } \sum_{j=1}^{R} q_j = 1, \tag{2.9}$$

using the Schwarz's inequality (see Theorem 3.6 in [LP17] for a detailed proof). The strong error assumption (SE_β) allows us to upper bound $\text{Var}(Y_h)$ and $\text{Var}(Z_j)$ by $\text{Var}(Y_0)\big(1+\theta h^{\beta/2}\big)^2$ with $\theta = \sqrt{\frac{V_1}{\text{Var}(Y_0)}}$ and $V_1\big(1+M^{\frac{\beta}{2}}\big)^2$ respectively. On the other hand, we assume that $\text{Cost}(Z_j) = \frac{(1+M^{-1})}{h} M^{j-1}$. Plugging theses estimates in (2.9) we obtain the *optimal* allocation policy used in this paper and given in Tables 2.1 and 2.2. Notice that this particular choice for the q_j is not unique, if we change (SE_β) with a different strong error assumption, for example with the sharp version, then we have to replace the upper bound for $\text{Var}(Z_j)$ with $V_1(M-1)^\beta$ and a new expression for the q_j follows. In the same spirit, the $\text{Cost}(Z_j)$ can be different and hence have an impact on the q_j, see [AGJC16] or the *nested* Monte Carlo methods as examples of alternative costs.

The second step is to select $h(\varepsilon) \in \mathcal{H}$ and $R(\varepsilon) \geqslant 2$ to minimize the cost of the *optimally allocated* estimator given a prescribed RMSE $\varepsilon > 0$. To do this we use the weak error assumption ($WE_{\alpha,\bar{R}}$) and we obtain

$$h(\varepsilon) = \mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha)^{\frac{1}{2\alpha}} |c_1|^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha}} M^{-(R-1)} \right\rceil$$

with c_1 the first coefficient in the weak error expansion, for the MLMC estimator. For the ML2R estimator we made the additional assumption $\tilde{c}_\infty = \lim_{R\to\infty} |c_R|^{\frac{1}{R}} \in (0, +\infty)$ and then we obtain

$$h(\varepsilon) = \mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha R)^{\frac{1}{2\alpha R}} \tilde{c}_\infty^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil.$$

The depth parameter $R \geqslant 2$ follows and the choice of N is directly related to the constraint (2.7).

We report in Tables 2.1 and 2.2 the ML2R and MLMC values for $R(\varepsilon)$, $h(\varepsilon)$, $q(\varepsilon) = (q_1(\varepsilon), \ldots, q_R(\varepsilon))$, $N(\varepsilon)$ computed in [LP17] and used throughout this paper. Note that these parameters are used in the web application of the LPMA at the address

http://simulations.lpma-paris.fr/multilevel. The following constants are used in this paper and in the Tables 2.1 and 2.2

$$\theta = \sqrt{\frac{V_1}{\mathrm{Var}(Y_0)}} \quad \text{and} \quad \widetilde{c}_\infty = \lim_{R\to\infty} |c_R|^{\frac{1}{R}} \in (0, +\infty)$$

and

$$\underline{C}_{M,\beta} = \frac{1 + M^{\frac{\beta}{2}}}{\sqrt{1 + M^{-1}}} \quad \text{and} \quad \bar{C}_{M,\beta} = \left(1 + M^{\frac{\beta}{2}}\right)\sqrt{1 + M^{-1}}.$$

Notice that $1+M^{\frac{\beta}{2}}$ comes from the (SE_β) and $\sqrt{1+M^{-1}}$ from the cost, hence the constants $\underline{C}_{M,\beta}$ and $\bar{C}_{M,\beta}$ depend on them, but on anything else.

$R(\varepsilon)$	$\left\lceil \frac{1}{2} + \frac{\log\left(\widetilde{c}_\infty^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log\left(\widetilde{c}_\infty^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)}\right)^2 + 2\frac{\log\left(\sqrt{1+4\alpha}/\varepsilon\right)}{\alpha\log(M)}} \right\rceil$
$h(\varepsilon)$	$\mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha R)^{\frac{1}{2\alpha R}} \widetilde{c}_\infty^{\frac{1}{\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil$
$q(\varepsilon)$	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^*\theta h^{\frac{\beta}{2}} \underline{C}_{M,\beta} \lvert \mathbf{W}_j(R,M)\rvert M^{-\frac{1+\beta}{2}(j-1)},\ j=2,\dots,R;\ \sum_{1\leqslant j\leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1+\frac{1}{2\alpha R}\right) \frac{\mathrm{Var}(Y_0)\left(1 + \theta h^{\frac{\beta}{2}} + \theta h^{\frac{\beta}{2}} \bar{C}_{M,\beta} \sum_{j=2}^{R} \lvert \mathbf{W}_j(R,M)\rvert M^{\frac{1-\beta}{2}(j-1)}\right)}{\varepsilon^2 \mu^*}$

TABLE 2.1 – Optimal parameters for the ML2R estimator.

$R(\varepsilon)$	$\left\lceil 1 + \frac{\log\left(\lvert c_1\rvert^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} + \frac{\log(\sqrt{1+2\alpha}/\varepsilon)}{\alpha\log(M)} \right\rceil$
$h(\varepsilon)$	$\mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha)^{\frac{1}{2\alpha}} \lvert c_1\rvert^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha}} M^{-(R-1)} \right\rceil$
$q(\varepsilon)$	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^*\theta h^{\frac{\beta}{2}} \underline{C}_{M,\beta} M^{-\frac{1+\beta}{2}(j-1)},\ j=2,\dots,R;\ \sum_{1\leqslant j\leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1+\frac{1}{2\alpha}\right) \frac{\mathrm{Var}(Y_0)\left(1 + \theta h^{\frac{\beta}{2}} + \theta h^{\frac{\beta}{2}} \bar{C}_{M,\beta} \sum_{j=2}^{R} M^{\frac{1-\beta}{2}(j-1)}\right)}{\varepsilon^2 \mu^*}$

TABLE 2.2 – Optimal parameters for the MLMC estimator.

In what follows, we will shorter these notations by setting

$$R(\varepsilon) = \left\lceil C_R^{(1)} + \sqrt{C_R^{(2)} + \frac{2}{\alpha\log(M)}\log\left(\frac{1}{\varepsilon}\right)} \right\rceil \tag{2.10}$$

with

$$C_R^{(1)} = \frac{1}{2} + \frac{\log\left(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} \quad \text{and} \quad C_R^{(2)} = \left(\frac{1}{2} + \frac{\log\left(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)}\right)^2 + 2\frac{\log\left(\sqrt{1+4\alpha}\right)}{\alpha\log(M)}$$

for ML2R and

$$R(\varepsilon) = \left\lceil C_R^{(1)} + \frac{1}{\alpha\log(M)}\log\left(\frac{1}{\varepsilon}\right)\right\rceil \tag{2.11}$$

with

$$C_R^{(1)} = 1 + \frac{\log\left(|c_1|^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} + \frac{\log\left(\sqrt{1+2\alpha}\right)}{\alpha\log(M)}$$

for MLMC.

2.3 Main results

The asymptotic behaviour, as ε goes to 0, of the parameters given in Tables 2.1 and 2.2 will be exposed in Section 2.4. We proceed here to the analysis of the asymptotic behaviour of the estimator $I_\pi^N(\varepsilon) := I_{\pi(\varepsilon)}^{N(\varepsilon)}$ as $\varepsilon \to 0$.

2.3.1 Strong Law of Large Numbers

We will first prove a Strong Law of Large Numbers, namely

Theorem 2.3.1 (Strong Law of Large Numbers)**.** *Let $p \geq 2$. Assume $(WE_{\alpha,\bar{R}})$ for all $\bar{R} \geq 1$ and $Y_0 \in L^p$. Assume furthermore the following L^p–strong error rate assumption*

$$\exists\, \beta > 0, V_1^{(p)} \geqslant 0, \quad \left\|Y_h - Y_0\right\|_p^p = \mathbf{E}\left[\left|Y_h - Y_0\right|^p\right] \leqslant V_1^{(p)} h^{\frac{\beta}{2}p}, \quad h \in \mathcal{H}\,. \tag{2.12}$$

Then, for every sequence of positive real numbers $(\varepsilon_k)_{k\geq 1}$ such that $\sum_{k\geq 1}\varepsilon_k^p < +\infty$, both MLMC and ML2R estimators satisfy

$$I_\pi^N(\varepsilon_k) \xrightarrow{a.s.} I_0, \quad \text{as } k \to +\infty. \tag{2.13}$$

2.3.2 Central Limit Theorems

A necessary condition for a Central Limit Theorem to hold will be that the ratio between the variance of the estimator and ε converges as $\varepsilon \to 0$. It seems intuitive that (SE_β) should be reinforced by a sharper estimate as $h \to 0$. We define

$$Z(h) := \left(\frac{h}{M}\right)^{-\frac{\beta}{2}}\left(Y_{\frac{h}{M}} - Y_h\right) \quad \text{and} \quad Z_j := Z\left(\frac{\mathbf{h}}{n_{j-1}}\right). \tag{2.14}$$

A necessary condition to obtain a CLT is to assume that $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable. We state two results, the first one in the case $\beta > 1$ and the second one in the case $\beta \leqslant 1$.

2.3.2.1 Case $\beta > 1$

In this case, note that following (SE_β) we have $\sup_{j\geqslant 1} \mathrm{Var}(Z_j) \leqslant V_1 \left(1 + M^{\frac{\beta}{2}}\right)^2$.

Theorem 2.3.2 (Central Limit Theorem, $\beta > 1$)**.** *Assume (SE_β) for $\beta > 1$ and that $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable. We set*

$$\sigma_1^2 = \frac{1}{\Sigma}\frac{\mathrm{Var}(Y_{\mathbf{h}})}{\mathrm{Var}(Y_0)(1+\theta\mathbf{h}^{\frac{\beta}{2}})} \quad \text{and} \quad \sigma_2^2 = \frac{1}{\Sigma}\frac{\mathbf{h}^{\frac{\beta}{2}}\sum_{j\geq 2} M^{\frac{1-\beta}{2}(j-1)}\mathrm{Var}(Z_j)}{\sqrt{\mathrm{Var}(Y_0)V_1}\underline{C}_{M,\beta}} \tag{2.15}$$

with

$$\Sigma = \Sigma(M,\beta,\theta,\mathbf{h}) = \left[1 + \theta\mathbf{h}^{\frac{\beta}{2}}\left(1 + \bar{C}_{M,\beta}\frac{M^{\frac{1-\beta}{2}}}{1-M^{\frac{1-\beta}{2}}}\right)\right].$$

Then the following statements hold.

(a) ML2R estimator : Assume $(WE_{\alpha,\bar{R}})$ for all $\bar{R} \geq 1$. Then

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_1^2 + \sigma_2^2\big), \quad \text{as } \varepsilon \to 0. \tag{2.16}$$

(b) MLMC estimator : Assume $(WE_{\alpha,\bar{R}})$ for $\bar{R} = 1$. Then there exists, for every $\varepsilon > 0$, $m(\varepsilon)$ such that $\frac{M^{-\alpha}}{\sqrt{1+2\alpha}} \leqslant |m(\varepsilon)| \leqslant \frac{1}{\sqrt{1+2\alpha}}$ and

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} - m(\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2\alpha}{2\alpha+1}\left(\sigma_1^2 + \sigma_2^2\right)\right), \quad \text{as } \varepsilon \to 0. \tag{2.17}$$

Note that the variance of the first term $Y_{\mathbf{h}}$ associated to the coarse level contributes to the asymptotic variance of the estimator throughout σ_1^2, while the variances of the correcting levels, $\mathrm{Var}(Z_j), j \geq 2$, contribute throughout σ_2^2. The normalized bias of the ML2R estimator goes to 0 as $\varepsilon \to 0$, whereas the MLMC estimator has an *a priori* non-vanishing normalized bias term. This gain for ML2R is balanced by the variance, which is reduced of a factor $\frac{2\alpha}{1+2\alpha}$ for MLMC. The constraint (2.7) yields $\sigma_1^2 + \sigma_2^2 \leq 1$, which is easy to verify if we recall that

$$\mathrm{Var}(Y_{\mathbf{h}}) \leq \mathrm{Var}(Y_0)\left(1 + \theta\mathbf{h}^{\frac{\beta}{2}}\right)^2, \quad \mathrm{Var}(Z_j) \leq V_1\left(1 + M^{\frac{\beta}{2}}\right)^2 \quad \text{and} \quad m(\varepsilon)^2 \leq \frac{1}{1+2\alpha}.$$

2.3.2.2 Case $\beta \in (0, 1]$

In this case, we make the additional sharper assumption that $\lim_{h\to 0} \big\|Z(h)\big\|_2^2 = v_\infty(M,\beta)$. This assumption allows us to identify $\lim_{j\to+\infty} \mathrm{Var}(Z_j)$. More precisely, note that owing to the consistence of the strong and weak error $2\alpha \geqslant \beta$ and owing to $(WE_{\alpha,\bar{R}})$ we have

$$\mathbf{E}\left[Z_j\right] = \left(\frac{\mathbf{h}}{n_j}\right)^{-\frac{\beta}{2}} \mathbf{E}\left[Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right] = c_1(1-M^\alpha)\left(\frac{\mathbf{h}}{n_j}\right)^{\alpha-\frac{\beta}{2}} + o\left(\left(\frac{\mathbf{h}}{n_j}\right)^{\alpha-\frac{\beta}{2}}\right),$$

so that

$$\mathrm{Var}(Z_j) = \left\|Z\left(\frac{\mathbf{h}}{n_{j-1}}\right)\right\|_2^2 - c_1^2\left(1-M^\alpha\right)^2\left(\frac{\mathbf{h}}{n_j}\right)^{2\alpha-\beta} + o\left(\left(\frac{\mathbf{h}}{n_j}\right)^{2\alpha-\beta}\right).$$

We conclude that

$$\lim_{j\to+\infty} \mathrm{Var}(Z_j) = \begin{cases} v_\infty(M,\beta) & \text{if } 2\alpha > \beta, \\ v_\infty(M,\beta) - c_1^2(1-M^{\frac{\beta}{2}})^2 & \text{if } 2\alpha = \beta. \end{cases} \tag{2.18}$$

Theorem 2.3.3 (Central Limit Theorem, $0 < \beta \leqslant 1$)**.** *Assume* (SE_β) *for* $\beta \in (0,1]$. *Assume that* $\big(Z(h)\big)_{h\in\mathcal{H}}$ *is* L^2*–uniformly integrable and assume furthermore* $\lim_{h\to 0} \big\|Z(h)\big\|_2^2 = v_\infty(M,\beta)$. *We set*

$$\sigma^2 = \begin{cases} v_\infty(M,\beta)\left(1+M^{\frac{\beta}{2}}\right)^{-2} V_1^{-1} & \text{if } 2\alpha > \beta, \\ \left(v_\infty(M,\beta) - c_1^2(1-M^{\frac{\beta}{2}})^2\right)\left(1+M^{\frac{\beta}{2}}\right)^{-2} V_1^{-1} & \text{if } 2\alpha = \beta. \end{cases} \tag{2.19}$$

Then the following statements hold.

(*a*) *ML2R estimator : Assume* ($WE_{\alpha,\bar{R}}$) *for all* $\bar{R} \geq 1$. *Then*

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}\big(0,\sigma^2\big), \quad \text{as } \varepsilon \to 0. \tag{2.20}$$

(*b*) *MLMC estimator : Assume* ($WE_{\alpha,\bar{R}}$) *for* $\bar{R} = 1$ *and that* $2\alpha > \beta$ *when* $\beta < 1$. *Then there exists for every* $\varepsilon > 0$, $m(\varepsilon)$ *such that* $\frac{M^{-\alpha}}{\sqrt{1+2\alpha}} \leqslant \big|m(\varepsilon)\big| \leqslant \frac{1}{\sqrt{1+2\alpha}}$ *and*

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} - m(\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2\alpha}{2\alpha+1}\sigma^2\right), \quad \text{as } \varepsilon \to 0. \tag{2.21}$$

We will see in the proof that the asymptotic variance corresponds to the variance associated to the correcting levels.

2.3.3 Practitioner's corner

In the proof of Theorems 2.3.2 and 2.3.3 we will obtain the more precise expansion

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} = m(\varepsilon) + \Sigma_2\zeta_2^\varepsilon + \frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\Sigma_1\zeta_1^\varepsilon, \quad \text{as } \varepsilon \to 0,$$

where ζ_1^ε and ζ_2^ε are two independent variables such that $(\zeta_1^\varepsilon, \zeta_2^\varepsilon) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_2)$ as $\varepsilon \to 0$, and the real values Σ_1 and Σ_2 depend on whether we are in the MLMC or in the ML2R case and on the value of β. Fundamentally Σ_1 comes from the variance of the first coarse level and Σ_2 from the sum of variances of the correcting levels.

When $\beta > 1$, we will prove in Lemma 2.4.3 that $\varepsilon\sqrt{N(\varepsilon)}$ converges to a constant as $\varepsilon \to 0$, hence both the coarse and the refined levels contribute to the asymptotic of the estimator.

When $\beta \leq 1$, we will see that $\left(\varepsilon\sqrt{N(\varepsilon)}\right)^{-1} \to 0$ as $\varepsilon \to 0$ so that, asymptotically, the variance of the coarse level fades and only the refined levels contribute to the asymptotic variance. Still, it is commonly known in the Multilevel framework that the coarse level is the one with the biggest size (speaking in terms of N_j), hence this term is not really negligible. We can go through this contradiction by observing the *inverse convergence rate to* 0, namely $\varepsilon\sqrt{N(\varepsilon)}$. It is equivalent, up to a constant, to $\sqrt{R(\varepsilon)}$ when $\beta = 1$ and $M^{\frac{1-\beta}{4}R(\varepsilon)}$ when $\beta < 1$.

- For ML2R, owing to the expression of $R(\varepsilon)$ given in (2.10), $\varepsilon\sqrt{N(\varepsilon)} \sim C\big(\log(1/\varepsilon)\big)^{\frac{1}{4}}$ where C is a positive constant when $\beta = 1$ and $\varepsilon\sqrt{N(\varepsilon)} = o\left(\varepsilon^{-\eta}\right)$ for all $\eta > 0$ when $\beta < 1$. Hence the convergence rate to 0 of $\left(\varepsilon\sqrt{N(\varepsilon)}\right)^{-1}$ is very slow. By contrast, $\Sigma_1 \gg \Sigma_2$, since Σ_1 is related to the variance of the coarse level which roughly approximates the value of interest whereas Σ_2 is related to the variance of the refined levels supposed to be smaller a priori. Hence the product $\left(\varepsilon\sqrt{N(\varepsilon)}\right)^{-1}\Sigma_1$ turns out not to be negligible with respect to Σ_2 for the values of the RMSE ε usually prescribed in applications.

- For MLMC, we get $\varepsilon\sqrt{N(\varepsilon)} \sim C\sqrt{\log(1/\varepsilon)}$, C positive constant, for $\beta = 1$ and $\varepsilon\sqrt{N(\varepsilon)} \sim C'\varepsilon^{-\frac{1-\beta}{4\alpha}}$ for $\beta < 1$. Hence, when $\beta < 1$, the slow convergence phenomenon is still observed though less significant.

▷ *Impact of the weights $\mathbf{W}_j^R$, $j = 1, \dots, R$ on the asymptotic behaviour of the ML2R estimator :* When $\beta \geq 1$, one observes that neither the rate of convergence nor the asymptotic variance of the estimator depends in any way upon the weights $\mathbf{W}_j^R$, $j = 1, \dots, R$. If $\beta < 1$ it depends in a somewhat hidden way through the multiplicative constant of $\varepsilon^{-2}M^{\frac{1-\beta}{2}R(\varepsilon)}$ in the asymptotic of $N(\varepsilon)$ (see Lemma 2.4.3 for more details). However, at finite range, it may have an impact on the variance of the estimator, having however in mind that, by construction, the depth of the ML2R estimator is lower than that of the MLMC which tempers this effect.

2.4 Auxiliary results

This Section contains some useful results for the proof of the Strong Law of Large Numbers and of the Central Limit Theorem. More in detail, we investigate the asymptotic behaviour as $\varepsilon \to 0$ of the optimal parameters given in Tables 2.1 and 2.2 and of the bias of the estimators and we analyze the weights of the ML2R estimator.

2.4.1 Asymptotic of the bias parameter and of the depth

An important property of MLMC and ML2R estimators is that $h(\varepsilon) \to \mathbf{h}$ and $R(\varepsilon) \to \infty$ as $\varepsilon \to 0$. The saturation of the bias parameter $\mathbf{h}$ is not intuitively obvious, indeed it is well known that $h(\varepsilon) \to 0$ as $\varepsilon \to 0$ for Crude Monte Carlo estimator. Still, this is a good property, because $h = \mathbf{h}$ is the choice which minimizes the cost of simulation of the variable Y_h, which we recall is inverse linear with respect to h. First of all, we retrace the computations that led to the choice of the optimal $h^*(\varepsilon)$ and $R^*(\varepsilon)$, starting from ML2R estimator. We define

$$h(\varepsilon, R) = (1 + 2\alpha R)^{-\frac{1}{2\alpha R}} |c_R|^{-\frac{1}{\alpha R}} \varepsilon^{\frac{1}{\alpha R}} M^{\frac{R-1}{2}}$$

and we recall that this is the optimized bias found in [LP17] at R fixed. Since the value of c_R is unknown, it is necessary to make the assumption $|c_R|^{\frac{1}{R}} \to \widetilde{c}$ as $R \to +\infty$ and $|c_R|^{-\frac{1}{\alpha R}}$ is replaced by $\widetilde{c}^{-\frac{1}{\alpha}}$. The value of $\widetilde{c}$ is also unknown and in the simulations we have to take an estimate of $\widetilde{c}$, that we write $\hat{c}$. We follow the lines of [LP17] and define the polynomial

$$P(R) = \frac{R(R-1)}{2}\log(M) - R\log(K) - \frac{1}{\alpha}\log\left(\frac{\sqrt{1+4\alpha}}{\varepsilon}\right) \tag{2.22}$$

where $K = \hat{c}^{\frac{1}{\alpha}}\mathbf{h}$. We set $R_+(\varepsilon)$ the positive zero of $P(R)$. The optimal value for the depth of the ML2R estimator is $R^*(\varepsilon) = \lceil R_+(\varepsilon) \rceil$. We notice that $P(R^*(\varepsilon)) \geq 0$, $R^*(\varepsilon) \to +\infty$ as $\varepsilon \to 0$, and R^* is increasing in $\hat{c}$. We can rewrite

$$h(\varepsilon, R) = \left(\frac{1+4\alpha}{1+2\alpha R}\right)^{\frac{1}{2\alpha R}} \left(\frac{\hat{c}}{|c_R|^{\frac{1}{R}}}\right)^{\frac{1}{\alpha}} e^{\frac{P(R)}{R}}\mathbf{h}.$$

We notice that

$$h(\varepsilon, R_+) = \left(\frac{1+4\alpha}{1+2\alpha R_+}\right)^{\frac{1}{2\alpha R_+}} \left(\frac{\hat{c}}{|c_{R_+}|^{\frac{1}{R_+}}}\right)^{\frac{1}{\alpha}} \mathbf{h}.$$

The optimal choice for the bias is the projection of $h(\varepsilon, R^*(\varepsilon))$ on the set $\mathcal{H} = \left\{\frac{\mathbf{h}}{n} : n \in \mathbf{N}\right\}$, which reads

$$h^*(\varepsilon) = \mathbf{h}\left\lceil \frac{\mathbf{h}}{h(\varepsilon, R^*(\varepsilon))} \right\rceil^{-1}.$$

When we replace $|c_R|^{\frac{1}{R}}$ with $\hat{c}$, we finally obtain

$$h^*(\varepsilon) = \frac{\mathbf{h}}{\left\lceil \mathbf{h}(1+2\alpha R^*)^{\frac{1}{2\alpha R^*}} \hat{c}^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha R^*}} M^{-\frac{R^*-1}{2}} \right\rceil} = \mathbf{h}\left\lceil \frac{\mathbf{h}}{h(\varepsilon, R^*)\left(|c_{R^*}|^{\frac{1}{R^*}}\hat{c}^{-1}\right)^{\frac{1}{\alpha}}} \right\rceil^{-1}.$$

Let us analyze the denominator

$$\frac{\mathbf{h}}{h(\varepsilon, R^*)\left(|c_{R^*}|^{\frac{1}{R^*}}\hat{c}^{-1}\right)^{\frac{1}{\alpha}}} = \left(\frac{1+4\alpha}{1+2\alpha R^*}\right)^{-\frac{1}{2\alpha R^*}} e^{-\frac{P(R^*)}{R^*}}.$$

Since $P(R^*) \geq 0$ and since for R large enough the function $\left(\frac{1+4\alpha}{1+2\alpha R}\right)^{-\frac{1}{2\alpha R}} \nearrow 1$, hence, up to reducing $\bar{\varepsilon}$,

$$\left(\frac{1+4\alpha}{1+2\alpha R^*}\right)^{-\frac{1}{2\alpha R^*}} e^{-\frac{P(R^*)}{R^*}} \leq 1, \quad \text{for all } \varepsilon \in (0, \bar{\varepsilon}), \tag{2.23}$$

which yields

$$\left\lceil \frac{\mathbf{h}}{h(\varepsilon, R^*)\left(|c_{R^*}|^{\frac{1}{R^*}}\hat{c}^{-1}\right)^{\frac{1}{\alpha}}} \right\rceil = 1 \quad \text{and} \quad h^*(\varepsilon) = \mathbf{h}$$

.

For MLMC we may follow the same reasoning starting from $h(\varepsilon, R) = (1 + 2\alpha)^{-\frac{1}{2\alpha}}|c_1|^{-\frac{1}{\alpha}}\varepsilon^{\frac{1}{\alpha}}M^{R-1}$. We just showed the following

Proposition 2.4.1. *There exists $\bar{\varepsilon} > 0$ such that $h^*(\varepsilon) = \mathbf{h}$ for all $\varepsilon \in (0, \bar{\varepsilon}]$.*

In what follows, we will always assume that $\varepsilon \in (0, \bar{\varepsilon}]$ and $h^*(\varepsilon) = \mathbf{h}$. This threshold $\bar{\varepsilon}$ can be reduced in what follows line to line.

As $\varepsilon \to 0$, we have $R = R^*(\varepsilon) \to +\infty$ at the rate $\sqrt{\frac{2}{\alpha \log(M)} \log\left(\frac{1}{\varepsilon}\right)}$ in the ML2R case and $\frac{1}{\alpha \log(M)} \log\left(\frac{1}{\varepsilon}\right)$ in the MLMC case.

2.4.2 Asymptotic of the bias and robustness

As part of a Central Limit Theorem, we will be faced to the quantity $\frac{\mu(\mathbf{h}, R(\varepsilon), M)}{\varepsilon}$, where

$$\mu(\mathbf{h}, R(\varepsilon), M) = \mathbf{E}\left[I_\pi^N(\varepsilon)\right] - I_0$$

is the bias of the estimator. This leads us to analyze carefully its asymptotic behavior as $\varepsilon \to 0$. Under the $(WE_{\alpha,\bar{R}})$ assumption, the bias of a Crude Monte Carlo estimator reads

$$\mu(h) = c_1 h^\alpha (1 + \eta_1(h)), \quad \lim_{h\to 0} \eta_1(h) = 0.$$

The bias of Multilevel estimators is dramatically reduced compared to the Crude Monte Carlo, more precisely the following Proposition is proved in [LP17] :

Proposition 2.4.2. *The following statements hold.*

(a) MLMC : Assume $(WE_{\alpha,\bar{R}})$ with $\bar{R} = 1$.

$$\mu(h, R, M) = c_1 \left(\frac{h}{M^{R-1}}\right)^\alpha \left(1 + \eta_1\left(\frac{h}{M^{R-1}}\right)\right) \tag{2.24}$$

with $\lim_{h\to 0} \eta_1(h) = 0$.

(b) ML2R : Assume $(WE_{\alpha,\bar{R}})$ for all $\bar{R} \geq 1$.

$$\mu(h, R, M) = (-1)^{R-1} c_R \left(\frac{h^R}{M^{\frac{R(R-1)}{2}}}\right)^\alpha (1 + \eta_{R,\underline{n}}(h)) \tag{2.25}$$

where

$$\eta_{R,\underline{n}}(h) = (-1)^{R-1} M^{\alpha\frac{R(R-1)}{2}} \sum_{r=1}^{R} \frac{w_r}{n_r^{\alpha R}} \eta_R(\frac{h}{n_R})$$

with $\lim_{h\to 0} \eta_R(h) = 0$.

We notice that the ML2R estimator requires and takes full advantage of a higher order of the expansion of the bias error $(WE_{\alpha,\bar{R}})$, whereas the MLMC estimator only needs a first order expansion. As the computations were made under the constraint $\left\|I_\pi^N - I_0\right\|_2 \leq \varepsilon$, we have clearly that $\frac{|\mu(\mathbf{h},R(\varepsilon),M)|}{\varepsilon} \leq 1$. We focus our attention on the constants $\widetilde{c}_\infty$ and c_1, which a priori we do not know and that we replace in the simulations by $\hat{c}_\infty = \hat{c}_1 = 1$. If we plug the values of $h(\varepsilon) = \mathbf{h}$ and $R(\varepsilon)$ in the formulas for the bias, owing to (2.22) and (2.23) we get, for ML2R,

$$\begin{aligned}|\mu(\mathbf{h}, R(\varepsilon), M)| &= |c_{R(\varepsilon)}| \left(\frac{\mathbf{h}^{R(\varepsilon)}}{M^{R(\varepsilon)\frac{R(\varepsilon)-1}{2}}}\right)^\alpha = |c_{R(\varepsilon)}| \mathbf{h}^{\alpha R(\varepsilon)} \frac{1}{e^{\alpha P(R(\varepsilon))} K^{\alpha R(\varepsilon)}} \frac{\varepsilon}{\sqrt{1+4\alpha}} \\ &= \frac{|c_{R(\varepsilon)}|}{\hat{c}_\infty^{R(\varepsilon)}} \frac{1}{e^{\alpha P(R(\varepsilon))}} \frac{\varepsilon}{\sqrt{1+4\alpha}} \leq \frac{|c_{R(\varepsilon)}|}{\hat{c}_\infty^{R(\varepsilon)}} \frac{1}{\sqrt{1+2\alpha R(\varepsilon)}} \varepsilon\end{aligned}$$

and, for MLMC,

$$\left|\frac{c_1}{\hat{c}_1}\right| \frac{M^{-\alpha}}{\sqrt{1+2\alpha}} \varepsilon < |\mu(\mathbf{h}, R(\varepsilon), M)| \leq \left|\frac{c_1}{\hat{c}_1}\right| \frac{1}{\sqrt{1+2\alpha}} \varepsilon. \tag{2.26}$$

We set $m(\varepsilon) := \frac{|\mu(\mathbf{h},R(\varepsilon),M)|}{\varepsilon}$. Hence, when taking the true values $\hat{c}_\infty = \widetilde{c}_\infty$ and $\hat{c}_1 = c_1$, we get

$$\begin{cases} \lim_{\varepsilon\to 0} m(\varepsilon) = 0 & \text{for } ML2R, \\ \frac{M^{-\alpha}}{\sqrt{1+2\alpha}} < \lim_{\varepsilon\to 0} m(\varepsilon) \leq \frac{1}{\sqrt{1+2\alpha}} & \text{for } MLMC. \end{cases} \tag{2.27}$$

For ML2R estimators, if c_R has a polynomial growth depending on R, we have

$$\lim_{R\to+\infty} |c_R|^{\frac{1}{R}} = 1$$

and $\hat{c}_\infty = 1$ corresponds to the exact value of $\widetilde{c}_\infty$. If the growth of c_R is less than polynomial, the convergence to 0 in (2.27) still holds. The only uncertain case is when the growth of c_R is faster than polynomial. Then, if $\hat{c}_\infty \geq |c_R|^{\frac{1}{R}}$, $|\mu(\varepsilon)|/\varepsilon$ goes to 0 faster than $\frac{1}{\sqrt{1+2\alpha R(\varepsilon)}}$, but if we had taken $\hat{c}_\infty < 1$, we would have obtained

$$\lim_{R\to+\infty} \frac{|c_R|}{\hat{c}_\infty^R} = +\infty,$$

hence $\hat{c}_\infty < 1$ is definitely not a good choice. In conclusion, whenever the growth of c_R is at most polynomial, $\hat{c}_\infty = 1$ remains a good choice. When the growth is faster than polynomial it is better to overestimate $\hat{c}_\infty$ than to underestimate it. The remarkable fact is that, when we choose $\hat{c}_\infty$, we are not forced to have a very precise idea of the expression of c_R, but only of its growth rate. The choice of $\hat{c}_1$ for MLMC estimator is less robust, since it is obvious that if we overestimate c_1 the inequality $|\mu(\varepsilon)|/\varepsilon \leq 1/\sqrt{1+2\alpha}$ still holds, but if we underestimate it we eventually may not have $\frac{|\mu(\varepsilon)|}{\varepsilon} \leq 1$ as expected. Hence the bias for the MLMC estimator is very connected to an accurate enough estimation of c_1.

In Figures 2.1a and 2.1b we show the values of $|c_1|$ estimated with the formula

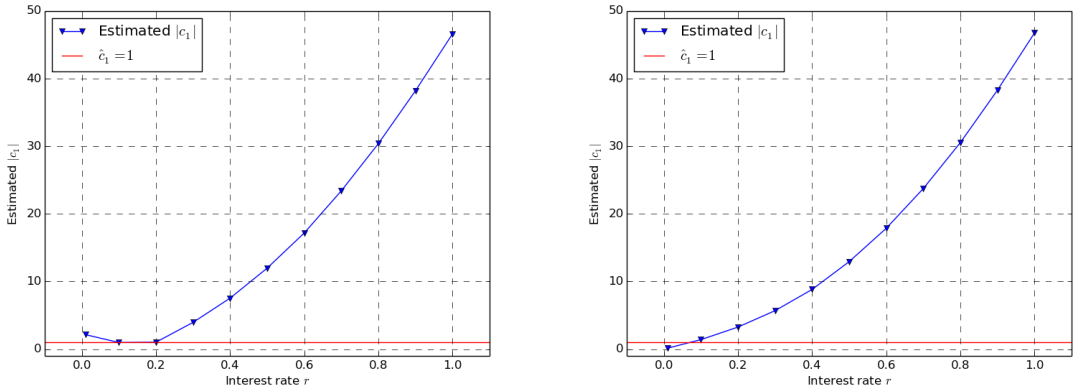
$$c_1 = \left(\frac{h}{2} - h\right)^{-1} \left(\mathbf{E}\left[Y_{\frac{h}{2}}\right] - \mathbf{E}\left[Y_h\right]\right)$$

compared to the value plugged in the simulations $\hat{c}_1 = 1$, for a Call option in a Black-Scholes model with $X_0 = 100$, $K = 80$, $T = 1$, $\sigma = 0.4$ and making the interest rate vary as follows $r = 0.01, 0.1, 0.2, \ldots, 0.9, 1$. We simulated $\mathbf{E}[Y_h]$, with $h = T/20$, using an Euler and a Milstein discretization scheme and making a Crude Monte Carlo simulation of size $N = 10^8$.

In Figures 2.2a and 2.2b we show the absolute value of the empirical bias for different values of r. In the simulations, we fixed $\hat{c}_1 = 1$ and $\hat{c}_\infty = 1$. We can observe that when $|c_1|$ is underestimated, the bias for MLMC and Crude Monte Carlo estimators do not satisfy the constraint $|\mu(\varepsilon)| \leq \varepsilon$, whereas the ML2R estimator appears to be less sensible to the estimation of $\tilde{c}$.

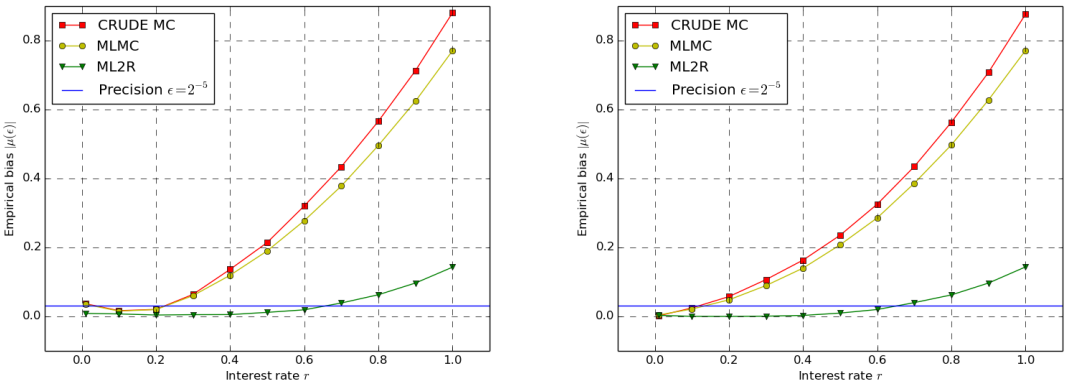
2.4.3 Properties of the weights of the ML2R estimator

One significant difficulty in the proof of the Central Limit Theorem that we stated in Theorems 2.3.2 and 2.3.3, is to deal with the weights $\mathbf{W}_j^R$ appearing in the ML2R estimator. Moreover, the analysis of the behaviour of the weights is necessary when studying the asymptotic of the parameters $q = (q_1, \ldots, q_R)$ and N. These weights are devised to kill the



(a) Euler scheme ($\alpha = 1, \beta = 1$) (b) Milstein scheme ($\alpha = 1, \beta = 2$).

FIGURE 2.1 – Estimated $|c_1| = \dfrac{\left|\mathbf{E}\left[Y_h\right] - \mathbf{E}\left[Y_{\frac{h}{2}}\right]\right|}{h - \frac{h}{2}}$ when r varies.



(a) Euler scheme ($\alpha = 1, \beta = 1$) (b) Milstein scheme ($\alpha = 1, \beta = 2$).

FIGURE 2.2 – Empirical bias $|\mu(\varepsilon)|$ for a Call option in a Black-Scholes model for a prescribed RMSE $\varepsilon = 2^{-5}$ and for different values of r, taking $\hat{c}_\infty = \hat{c}_1 = 1$.

coefficients $c_1, \dots, c_R$ in the bias expansion under the ($WE_{\alpha,\bar{R}}$). They are defined as

$$\mathbf{W}_j^R = \sum_{r=j}^{R} \mathbf{w}_r, \quad j = 1, \dots, R, \tag{2.28}$$

where the weights $\mathbf{w} = (\mathbf{w}_r)_{r=1,\dots,R}$ are the solution to the Vandermonde system $V\,\mathbf{w} = e_1$, the matrix V being defined by

$$V = V(1, n_2^{-\alpha}, \dots, n_R^{-\alpha}) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & n_2^{-\alpha} & \cdots & n_R^{-\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-1)} & \cdots & n_R^{-\alpha(R-1)} \end{pmatrix}. \tag{2.29}$$

Notice that $\mathbf{W}_1^R = 1$ by construction. In order to give a more tractable expression of the weights $\mathbf{W}_j^R$, one notices that the weights $\mathbf{w}$ admit a closed form given by Cramer's rule, namely

$$\mathbf{w}_\ell = a_\ell b_{R-\ell}, \quad \ell = 1, \dots, R,$$

where

$$a_\ell = \frac{1}{\prod_{1 \le k \le \ell-1}(1 - M^{-k\alpha})}, \quad \ell = 1, \dots, R,$$

with the convention $\prod_{k=1}^{0}(1 - M^{-k\alpha}) = 1$, and

$$b_\ell = (-1)^\ell \frac{M^{-\frac{\alpha}{2}\ell(\ell+1)}}{\prod_{1 \le k \le \ell}(1 - M^{-k\alpha})}, \quad \ell = 0, \dots, R.$$

As a consequence

$$\mathbf{W}_j^R = \sum_{\ell=j}^{R} a_\ell b_{R-\ell} = \sum_{\ell=0}^{R-j} a_{R-\ell} b_\ell, \quad j \in 1, \dots, R.$$

We will make an extensive use of the following properties, which are proved in Appendix A.

Lemma 2.4.1. *Let $\alpha > 0$ and the associated weights $(\mathbf{W}_j^R)_{j=1,\dots,R}$ given in (2.28).*

(a) $\lim_{\ell \to +\infty} a_\ell = a_\infty < +\infty$ *and* $\sum_{\ell=0}^{+\infty} |b_\ell| = \widetilde{B}_\infty < +\infty$.

(b) The weights $\mathbf{W}_j^R$ are uniformly bounded,

$$\forall R \in \mathbf{N}^*, \forall j \in \{1, \dots, R\}, \qquad \left|\mathbf{W}_j^R\right| \le a_\infty \widetilde{B}_\infty. \tag{2.30}$$

(c) For every $\gamma > 0$,

$$\lim_{R \to +\infty} \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{-\gamma(j-1)} = \frac{1}{M^\gamma - 1}.$$

(d) Let $\{v_j\}_{j \ge 1}$ be a bounded sequence of positive real numbers. Let $\gamma \in \mathbf{R}$ and assume that $\lim_{j \to +\infty} v_j = 1$ when $\gamma \ge 0$. Then the following limits hold :

$$\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} v_j \overset{R \to +\infty}{\sim} \begin{cases} \sum_{j \ge 2} M^{\gamma(j-1)} v_j < +\infty & \text{for} \quad \gamma < 0, \\ R & \text{for} \quad \gamma = 0, \\ M^{\gamma R} a_\infty \sum_{j \ge 1} \left|\sum_{\ell=0}^{j-1} b_\ell\right| M^{-\gamma j} & \text{for} \quad \gamma > 0. \end{cases}$$

2.4.4 Asymptotic of the allocation policy and of the size

Let us analyze the allocation policy $q = (q_1, \dots, q_R)$ for the ML2R case. Since

$$q_1(\varepsilon) = \mu^*(\varepsilon)\left(1 + \theta \mathbf{h}^{\frac{\beta}{2}}\right) \tag{2.31}$$

and

$$q_j(\varepsilon) = \theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta} \mu^*(\varepsilon) \left|\mathbf{W}_j^{R(\varepsilon)}\right| M^{-\frac{\beta+1}{2}(j-1)}, \quad j = 2, \dots, R(\varepsilon), \tag{2.32}$$

the condition $\sum_{j=1}^{R(\varepsilon)} q_j = 1$ yields

$$\mu^*(\varepsilon) = \left(1 + \theta \mathbf{h}^{\frac{\beta}{2}} \left(1 + \underline{C}_{M,\beta} \sum_{j=2}^{R(\varepsilon)} \left|\mathbf{W}_j^{R(\varepsilon)}\right| M^{-\frac{\beta+1}{2}(j-1)}\right)\right)^{-1}.$$

Owing to Lemma 2.4.1 (c) with $\gamma = \frac{\beta+1}{2}$, the limit of this term as $\varepsilon \to 0$ is

$$\boldsymbol{\mu}^* = \left(1 + \theta \mathbf{h}^{\frac{\beta}{2}} \left(1 + \frac{\underline{C}_{M,\beta}}{M^{\frac{1+\beta}{2}} - 1}\right)\right)^{-1}.$$

Moreover, for all $\varepsilon \in (0, \bar{\varepsilon}]$, the following inequalities hold :

$$\frac{1}{\bar{\mu}^*} := 1 + \theta \mathbf{h}^{\frac{\beta}{2}} \leq \frac{1}{\mu^*(\varepsilon)} \leq 1 + \theta \mathbf{h}^{\frac{\beta}{2}} \left(1 + \underline{C}_{M,\beta} a_\infty \widetilde{B}_\infty \frac{1}{1 - M^{-\frac{\beta+1}{2}}}\right) =: \frac{1}{\underline{\mu}^*}. \tag{2.33}$$

Remark 2.4.2. *If we set $\mathbf{W}_j^{R(\varepsilon)} = 1$ for all $j = 1, \dots, R(\varepsilon)$, and $a_\infty \widetilde{B}_\infty = 1$, we obtain the same results for the MLMC allocation policy.*

The asymptotic of the estimator size $N = N(\varepsilon)$ is given in the following Lemma.

Lemma 2.4.3. *We have $N = N(\varepsilon) \to +\infty$, as $\varepsilon \to 0$, with a convergence rate depending on β as follows :*
***Case** $\beta > 1$: We have $N(\varepsilon) \sim C_\beta \varepsilon^{-2}$ with*

$$C_\beta = \frac{\mathrm{Var}(Y_0)}{\boldsymbol{\mu}^*} \left[1 + \theta \mathbf{h}^{\frac{\beta}{2}} \left(1 + \bar{C}_{M,\beta} \frac{M^{\frac{1-\beta}{2}}}{1 - M^{\frac{1-\beta}{2}}}\right)\right] \begin{cases} 1 & \text{for ML2R,} \\ 1 + \dfrac{1}{2\alpha} & \text{for MLMC.} \end{cases} \tag{2.34}$$

***Case** $\beta \leq 1$: We recall the expression of $R(\varepsilon)$ given in (2.10) for ML2R and (2.11) for MLMC. Then*

$$N(\varepsilon) \sim C_\beta \varepsilon^{-2} \begin{cases} R(\varepsilon) & \text{if} \quad \beta = 1, \\ M^{\frac{1-\beta}{2} R(\varepsilon)} & \text{if} \quad \beta < 1 \end{cases}$$

where the constant C_β reads

$$C_\beta = \frac{\mathrm{Var}(Y_0)}{\boldsymbol{\mu}^*} \theta \mathbf{h}^{\frac{1}{2}} \bar{C}_{M,\beta} \begin{cases} 1 & \text{for ML2R,} \\ 1 + \dfrac{1}{2\alpha} & \text{for MLMC} \end{cases} \quad \text{if} \quad \beta = 1 \tag{2.35}$$

and

$$C_\beta = \frac{\mathrm{Var}(Y_0)}{\boldsymbol{\mu}^*}\theta\mathbf{h}^{\frac{\beta}{2}}\bar{C}_{M,\beta}\begin{cases} a_\infty \sum_{j\geq 1}\left|\sum_{\ell=0}^{j-1} b_\ell\right| M^{\frac{\beta-1}{2}j} & \text{for ML2R,} \\ \left(1+\frac{1}{2\alpha}\right)\frac{1}{M^{\frac{1-\beta}{2}}-1} & \text{for MLMC} \end{cases} \quad \text{if} \quad \beta < 1. \tag{2.36}$$

We notice that for $\beta \geq 1$ the asymptotic behaviour of $N(\varepsilon)$ for ML2R does not depend on the weights $\mathbf{W}_j^R$ and the difference between the coefficient C_β for ML2R and for MLMC estimator lies only in the factor $\left(1+\frac{1}{2\alpha}\right)$, whereas when $\beta < 1$ the asymptotic of the weights has an impact on the behaviour of $N(\varepsilon)$ for ML2R. Still, in this case we observe that if $a_\infty = 1$ and $\left|\sum_{\ell=0}^{j-1} b_\ell\right| = 1$ for all $j \geq 1$, then

$$a_\infty \sum_{j\geq 1}\left|\sum_{\ell=0}^{j-1} b_\ell\right| M^{\frac{\beta-1}{2}j} = \frac{1}{M^{\frac{1-\beta}{2}}-1}$$

and the factor $\left(1+\frac{1}{2\alpha}\right)$ appears again to be the only difference in the coefficient C_β of $N(\varepsilon)$ for the two estimators.

Démonstration. ▷ ML2R : The estimator size N reads

$$N = N(\varepsilon) = \left(1+\frac{1}{2\alpha R(\varepsilon)}\right)\frac{\mathrm{Var}(Y_0)}{\mu^*}\frac{1}{\varepsilon^2}\left(1+\theta\mathbf{h}^{\frac{\beta}{2}}+\theta\mathbf{h}^{\frac{\beta}{2}}\bar{C}_{M,\beta}\sum_{j=2}^{R(\varepsilon)}\left|\mathbf{W}_j^{R(\varepsilon)}\right| M^{\frac{1-\beta}{2}(j-1)}\right).$$

We notice that $R(\varepsilon) \to +\infty$ as $\varepsilon \to 0$ and use Lemma 2.4.1 (d) with $\gamma = \frac{1-\beta}{2}$, with $v_j = 1$ for each $j \geq 1$, to complete the proof on the ML2R framework.

▷ MLMC : The result follows directly from the convergence of the series $\sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)}$, since N reads

$$N = N(\varepsilon) = \left(1+\frac{1}{2\alpha}\right)\frac{\mathrm{Var}(Y_0)}{\mu^*}\frac{1}{\varepsilon^2}\left(1+\theta\mathbf{h}^{\frac{\beta}{2}}\left(1+\bar{C}_{M,\beta}\sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)}\right)\right),$$

as desired. □

2.5 Proofs

We will use the notations

$$\widetilde{I}_\varepsilon^1 := \frac{1}{N_1(\varepsilon)}\sum_{k=1}^{N_1(\varepsilon)} Y_{\mathbf{h}}^{(1),k} - \mathbf{E}\left[Y_{\mathbf{h}}^{(1),k}\right] \quad \text{and} \quad \widetilde{I}_\varepsilon^2 := \sum_{j=2}^{R(\varepsilon)}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\widetilde{Y}_j^k$$

where we set

$$\widetilde{Y}_j := Y_{\frac{\mathbf{h}}{n_j}}^{(j)} - Y_{\frac{\mathbf{h}}{n_{j-1}}}^{(j)} - \mathbf{E}\left[Y_{\frac{\mathbf{h}}{n_j}}^{(j)} - Y_{\frac{\mathbf{h}}{n_{j-1}}}^{(j)}\right], \quad j = 1,\ldots,R(\varepsilon). \tag{2.37}$$

These notations hold for both ML2R and MLMC estimators, where we set $\mathbf{W}_j^{R(\varepsilon)} = 1$, $j = 1, \dots, R(\varepsilon)$, for MLMC estimators. We notice that

$$I_\pi^N(\varepsilon) - I_0 = \widetilde{I}_\varepsilon^1 + \widetilde{I}_\varepsilon^2 + \mu(\mathbf{h}, R(\varepsilon), M) \tag{2.38}$$

where the bias $\mu(\mathbf{h}, R(\varepsilon), M) \to 0$ as $\varepsilon \to 0$ (see Section 2.4.2 for a detailed description of the bias).

2.5.1 Proof of Strong Law of Large Numbers

The proof of the Strong Law of Large Numbers is a consequence of the following Proposition.

Proposition 2.5.1. *Let $p \geq 2$. There exists a positive real constant $K(M, \beta, p)$ such that*

$$\mathbf{E}\left[\left|\widetilde{I}_\varepsilon^2\right|^p\right] \leq K(M, \beta, p)\varepsilon^p. \tag{2.39}$$

Proof. $\triangleright$ ML2R : We first give the proof of (2.39) for the ML2R estimator. As a first step we show that, for all $p \geq 2$,

$$\mathbf{E}\left[\left|\widetilde{Y}_j\right|^p\right] \leq C_{M,\beta,p} M^{-\frac{\beta p}{2}(j-1)}, \quad j = 1, \dots, R(\varepsilon), \tag{2.40}$$

with $C_{M,\beta,p} = 2^p V_1^{(p)} \left(1 + M^{\frac{\beta}{2}}\right)^p \mathbf{h}^{\frac{\beta p}{2}}$. By Minkowski's Inequality

$$\begin{aligned}\left(\mathbf{E}\left[\left|\widetilde{Y}_j\right|^p\right]\right)^{\frac{1}{p}} &\leq \left\|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right\|_p + \left|\mathbf{E}\left[Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right]\right| \\ &\leq \left\|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right\|_p + \left\|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right\|_1 \leq 2\left\|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\right\|_p.\end{aligned}$$

Applying again Minkowski's Inequality, the L^p-strong approximation assumption (2.12) yields (2.40).

As the random variables $\left(\widetilde{Y}_j^k\right)_{k\geq 1}$ are *i.i.d.* and the $(\widetilde{Y}_j)_{j=1,\dots,R(\varepsilon)}$ are centered and independent, Burkholder's Inequality (see [HH80], Theorem 2.10, p. 23) and (2.40) imply that there exists a positive universal real constant C_p such that

$$\begin{aligned}\mathbf{E}\left[\left|\widetilde{I}_\varepsilon^2\right|^p\right] &= \mathbf{E}\left[\left|\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)} \frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right|^p\right] \leq C_p\mathbf{E}\left[\left|\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\left(\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right)^2\right|^{\frac{p}{2}}\right] \\ &\leq C_p\left(\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\left\|\left(\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right)^2\right\|_{\frac{p}{2}}\right)^{\frac{p}{2}} = C_p\left(\sum_{j=2}^{R(\varepsilon)}\frac{|\,\mathbf{W}_j^{R(\varepsilon)}\,|^2}{N_j(\varepsilon)}\left(\mathbf{E}\left[\left|\widetilde{Y}_j\right|^p\right]\right)^{\frac{2}{p}}\right)^{\frac{p}{2}} \\ &\leq C_p C_{M,\beta,2}\left(\sum_{j=2}^{R(\varepsilon)}\frac{|\,\mathbf{W}_j^{R(\varepsilon)}\,|^2}{N_j(\varepsilon)}M^{-\beta(j-1)}\right)^{\frac{p}{2}}.\end{aligned}$$

As $N_j(\varepsilon) = \lceil N(\varepsilon) q_j(\varepsilon) \rceil \geq N(\varepsilon) q_j(\varepsilon)$, we derive that

$$\frac{1}{N_j(\varepsilon)} \leq \frac{1}{N(\varepsilon) q_j(\varepsilon)}, \quad j = 1, \dots, R(\varepsilon).$$

It follows from the expression of q_j given in (2.32) and from inequality (2.33) that

$$\frac{|\, \mathbf{W}_j^{R(\varepsilon)} \,|}{q_j(\varepsilon)} \leq \frac{1}{\theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta} \underline{\mu}^*} M^{\frac{\beta+1}{2}(j-1)}, \quad j = 2, \dots, R(\varepsilon).$$

Then, using that $\sup_{j \in \{1,\dots,R\}, R \geq 1} |\, \mathbf{W}_j^R \,| \leq a_\infty \widetilde{B}_\infty$, we get

$$\mathbf{E}\left[\left| \widetilde{I}_\varepsilon^2 \right|^p \right] \leq C_p C_{M,\beta,2} \left(a_\infty \widetilde{B}_\infty \right)^{\frac{p}{2}} \left(\theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta} \underline{\mu}^* \right)^{-\frac{p}{2}} \left(\frac{1}{N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)} \right)^{\frac{p}{2}}.$$

Owing to Lemma 2.4.3, up to reducing $\bar{\varepsilon}$, we have

$$\forall \varepsilon \in (0, \bar{\varepsilon}], \quad \frac{1}{N(\varepsilon)} \leq \frac{2}{C_\beta} \varepsilon^2 \begin{cases} 1 & \text{if } \beta > 1, \\ R(\varepsilon)^{-1} & \text{if } \beta = 1, \\ M^{-\frac{1-\beta}{2} R(\varepsilon)} & \text{if } \beta < 1. \end{cases} \tag{2.41}$$

Moreover,

$$\sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)} \leq \begin{cases} \frac{1}{1 - M^{\frac{1-\beta}{2}}} & \text{if } \beta > 1, \\ R(\varepsilon) & \text{if } \beta = 1, \\ \frac{M^{\frac{1-\beta}{2} R(\varepsilon)}}{M^{\frac{1-\beta}{2}} - 1} & \text{if } \beta < 1. \end{cases}$$

Then

$$\left(\frac{1}{N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)} \right)^{\frac{p}{2}} \leq K_1 \varepsilon^p$$

with

$$K_1 = K_1(M, \beta, p) = \left(\frac{2}{C_\beta} \right)^{\frac{p}{2}} \begin{cases} \left(1 - M^{\frac{1-\beta}{2}} \right)^{-\frac{p}{2}} & \text{if } \beta > 1, \\ 1 & \text{if } \beta = 1, \\ \left(M^{\frac{1-\beta}{2}} - 1 \right)^{-\frac{p}{2}} & \text{if } \beta < 1. \end{cases}$$

Hence (2.39) holds with

$$K(M, \beta, p) = C_p C_{M,\beta,2} \left(a_\infty \widetilde{B}_\infty \right)^{\frac{p}{2}} \left(\theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta} \underline{\mu}^* \right)^{-\frac{p}{2}} K_1.$$

▷ MLMC : The proof for the MLMC estimator follows the same steps, by replacing $\mathbf{W}_j^{R(\varepsilon)} = 1$, for $j = 1, \dots, R(\varepsilon)$, and $a_\infty \widetilde{B}_\infty = 1$. □

The Strong Law of Large Numbers follows as a consequence of Proposition 2.5.1.

Proof of Theorem 2.3.1. Owing to the decomposition (2.38), equation (2.13) amounts to proving

$$\widetilde{I}^1_{\varepsilon_k} \xrightarrow{a.s.} 0 \quad \text{as} \quad k \to +\infty \quad \text{and} \quad \widetilde{I}^2_{\varepsilon_k} \xrightarrow{a.s.} 0 \quad \text{as} \quad k \to +\infty. \tag{2.42}$$

As $\lim_{\varepsilon\to 0} N_1(\varepsilon) = +\infty$ and the $(Y^{(1),k}_{\mathbf{h}})_{k\geq 1}$ are *i.i.d.* and do not depend on ε, the convergence of $\widetilde{I}^1_{\varepsilon_k}$ is a direct application of the classical Strong Law of Large Numbers, for both ML2R and MLMC estimators.

To establish the *a.s.* convergence of $\widetilde{I}^2_{\varepsilon_k}$, owing to Proposition 2.5.1 it is straightforward that for all sequence of positive values $(\varepsilon_k)_{k\geq 1}$ such that $\varepsilon_k \to 0$ as $k \to +\infty$ and $\sum_{k\geq 1} \varepsilon_k^p < +\infty$

$$\sum_{k\geq 1} \mathbf{E}\left[\left|\widetilde{I}^2_{\varepsilon_k}\right|^p\right] < +\infty.$$

Hence, by Beppo-Levi's Theorem, $\sum_{k\geq 1}\left|\widetilde{I}^2_{\varepsilon_k}\right|^p < +\infty$ *a.s.*, which in turn implies $\widetilde{I}^2_{\varepsilon_k} \xrightarrow{a.s.} 0$ as $k \to +\infty$. □

2.5.2 Proof of the Central Limit Theorem

This subsection is devoted to the proof of Theorems 2.3.2 and 2.3.3. In order to satisfy a Lindeberg condition, we will need the assumption $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable. Owing to ($WE_{\alpha,\bar{R}}$), $\bar{R} = 1$,

$$\big|\mathbf{E}\left[Z(h)\right]\big| = \big|c_1(1 - M^\alpha)\big|h^{\alpha-\frac{\beta}{2}} + o(h^{\alpha-\frac{\beta}{2}}).$$

Since $2\alpha \geq \beta$, this deterministic sequence $\big(\mathbf{E}\left[Z(h)\right]\big)_{h\in\mathcal{H}}$ is bounded. Hence, the L^2–uniform integrability of $(Z(h))_{h\in\mathcal{H}}$ yields the L^2–uniform integrability of the centered sequence $\big(\widetilde{Z}(h)\big)_{h\in\mathcal{H}} = (Z(h) - \mathbf{E}\left[Z(h)\right])_{h\in\mathcal{H}}$.

One criterion to verify the L^2–uniform integrability is the following.

Lemma 2.5.1. *The following statements hold.*

(*a*) *If there exists a* $p > 2$ *such that* $\sup_{h\in\mathcal{H}} \big\|Z(h)\big\|_p < +\infty$ *the family* $\big(Z(h)\big)_{h\in\mathcal{H}}$ *is* L^2*–uniformly integrable.*

(*b*) *If there exists a random variable* $D^{(M)} \in L^2$ *such that, as* $h \to 0$,

$$Z(h) \xrightarrow{\mathcal{L}} D^{(M)}$$

then the following conditions are equivalent (see [Bil99], Theorem 3.6) :

(i) The family $\big(Z(h)\big)_{h\in\mathcal{H}}$ *is* L^2*–uniformly integrable.*

(ii) $\lim_{h\to 0} \big\|Z(h)\big\|_2 = \big\|D^{(M)}\big\|_2$.

Now we are in a position to prove the Central Limit Theorem, in both cases $\beta > 1$ and $\beta \in (0, 1]$.

Proof of Theorems 2.3.2 and 2.3.3. Owing to the decomposition (2.38) (with $\mathbf{W}_j^{R(\varepsilon)} = 1$, $j = 1, \ldots, R(\varepsilon)$ for MLMC estimator)

$$\frac{I^N_\pi(\varepsilon) - I_0}{\varepsilon} = \frac{\widetilde{I}^1_\varepsilon}{\varepsilon} + \frac{\widetilde{I}^2_\varepsilon}{\varepsilon} + \frac{\mu(\mathbf{h}, R(\varepsilon), M)}{\varepsilon}$$

where $\widetilde{I}_\varepsilon^1$ and $\widetilde{I}_\varepsilon^2$ are independent. The bias term has already been treated in (2.27).

$\triangleright$ ML2R : Formulas (2.16) and (2.20) amount to proving, as $\varepsilon \to 0$,

$$\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\mathrm{Var}(Y_{\mathbf{h}})}{\mathbf{q}_1}\right) \quad (2.43) \qquad \text{and} \qquad \frac{\widetilde{I}_\varepsilon^2}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_2^2\right) \quad (2.44)$$

with $\sigma_2 = \sigma$ for $\beta \in (0,1]$. Indeed, for (2.43) let us write $\dfrac{\widetilde{I}_\varepsilon^1}{\varepsilon} = \dfrac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1$. Using Lemma 2.4.3, $N(\varepsilon)$ reads

$$N(\varepsilon) \overset{\varepsilon\to 0}{\sim} C_\beta \varepsilon^{-2} \begin{cases} 1 & \text{if } \beta > 1, \\ R(\varepsilon) & \text{if } \beta = 1, \\ M^{\frac{1-\beta}{2}R(\varepsilon)} & \text{if } \beta < 1. \end{cases} \quad (ML2R)$$

In particular, since $R(\varepsilon) \to +\infty$ as $\varepsilon \to 0$, when $\beta \le 1$, $\dfrac{1}{\sqrt{N(\varepsilon)}} = o(\varepsilon)$ and the term $\dfrac{\widetilde{I}_\varepsilon^1}{\varepsilon} \to 0$ in probability. Since $Y_{\mathbf{h}}^{(1),k}$ does not depend on ε, $N_1(\varepsilon) \to +\infty$ and $N_1(\varepsilon)/N(\varepsilon) \to \mathbf{q}_1$ as $\varepsilon \to 0$, the asymptotic behaviour of the first term is driven by a regular Central Limit Theorem at rate $\sqrt{N(\varepsilon)}$, *i.e.*

$$\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 = \sqrt{N(\varepsilon)}\left[\frac{1}{N_1(\varepsilon)}\sum_{k=1}^{N_1(\varepsilon)}\left(Y_{\mathbf{h}}^{(1),k} - \mathbf{E}\left[Y_{\mathbf{h}}^{(1),k}\right]\right)\right] \xrightarrow{\varepsilon\to 0} \mathcal{N}\left(0, \frac{\mathrm{Var}\left(Y_{\mathbf{h}}\right)}{\mathbf{q}_1}\right)$$

which proves (2.43).

We will use Lindeberg's Theorem for triangular arrays of martingale increments (see Corollary 3.1 p.58 in [HH80]) to establish (2.44). The random variables $\widetilde{Y}_j^k$ being centered and independent, the variance reads

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right) &= \frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\left(\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\right)^2 N_j(\varepsilon)\,\mathrm{Var}\left(\widetilde{Y}_j\right) \\ &= \frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{N_j(\varepsilon)}\mathrm{Var}\left(\widetilde{Y}_j\right).\end{aligned}$$

Noticing that $0 \le \frac{1}{x} - \frac{1}{\lceil x\rceil} \le \frac{1}{x^2}$, $x > 0$, and that $N_j(\varepsilon) = \lceil q_j(\varepsilon)N(\varepsilon)\rceil$, we derive

$$\begin{aligned}&\left|\frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{N_j(\varepsilon)}\mathrm{Var}\left(\widetilde{Y}_j\right) - \frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{q_j(\varepsilon)N(\varepsilon)}\mathrm{Var}\left(\widetilde{Y}_j\right)\right| \\ &\qquad \le \frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{\left(q_j(\varepsilon)N(\varepsilon)\right)^2}\mathrm{Var}\left(\widetilde{Y}_j\right).\end{aligned}$$

The conclusion will follow from

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{N(\varepsilon)q_j(\varepsilon)}\,\mathrm{Var}\left(\widetilde{Y}_j\right)=\sigma_2^2 \tag{2.45}$$

and

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{(q_j(\varepsilon)N(\varepsilon))^2}\,\mathrm{Var}\left(\widetilde{Y}_j\right)=0. \tag{2.46}$$

Owing to the definition of Z_j given in (2.14), we get $\mathrm{Var}(\widetilde{Y}_j)=\left(\frac{\mathbf{h}}{n_j}\right)^{\beta}\mathrm{Var}(Z_j)$ and, using the expression of $q_j(\varepsilon)$ given in (2.32), we obtain

$$\frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{N(\varepsilon)q_j(\varepsilon)}\,\mathrm{Var}\left(\widetilde{Y}_j\right)=\frac{1}{\varepsilon^2 N(\varepsilon)}\frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta\underline{C}_{M,\beta}\mu^*(\varepsilon)}\sum_{j=2}^{R(\varepsilon)}\left|\mathbf{W}_j^{R(\varepsilon)}\right|M^{\frac{1-\beta}{2}(j-1)}\,\mathrm{Var}(Z_j).$$

• *Case* $\beta>1$: Owing to the expression of $N(\varepsilon)$ given in Lemma 2.4.3 when $\beta>1$,

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2 N(\varepsilon)}\frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta\underline{C}_{M,\beta}\mu^*(\varepsilon)}=\frac{1}{\Sigma}\frac{\mathbf{h}^{\frac{\beta}{2}}}{\sqrt{\mathrm{Var}(Y_0)V_1}\underline{C}_{M,\beta}}$$

and owing to the limit in Lemma 2.4.1 (d) with $\gamma=\frac{1-\beta}{2}<0$,

$$\sum_{j=2}^{R(\varepsilon)}\left|\mathbf{W}_j^{R(\varepsilon)}\right|M^{\frac{1-\beta}{2}(j-1)}\,\mathrm{Var}(Z_j)=\sum_{j=2}^{+\infty}M^{\frac{1-\beta}{2}(j-1)}\,\mathrm{Var}(Z_j)<+\infty.$$

Hence the convergence of the variance (2.45) holds for Theorem 2.3.2.
• *Case* $\beta\le 1$: Owing to the expression of $N(\varepsilon)$ given in Lemma 2.4.3 when $\beta\le 1$, we get, as $\varepsilon\to 0$,

$$\frac{1}{\varepsilon^2 N(\varepsilon)}\frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta\underline{C}_{M,\beta}\mu^*(\varepsilon)}$$
$$\sim\frac{1}{V_1\left(1+M^{\frac{\beta}{2}}\right)^2}\begin{cases}(R(\varepsilon))^{-1} & \text{if } \beta=1,\\ \left(M^{\frac{1-\beta}{2}R(\varepsilon)}a_\infty\sum_{j\ge 1}\left|\sum_{\ell=0}^{j-1}b_\ell\right|M^{\frac{\beta-1}{2}j}\right)^{-1} & \text{if } \beta<1.\end{cases}$$

We notice that $\lim_{j\to+\infty}\mathrm{Var}(Z_j)=\tilde{v}_\infty(M,\beta)$ if $2\alpha>\beta$ and $\lim_{j\to+\infty}\mathrm{Var}(Z_j)=\tilde{v}_\infty(M,\beta)-c_1^2\left(1-M^{\frac{\beta}{2}}\right)^2$ if $2\alpha=\beta$. Hence, owing to the limit in Lemma 2.4.1 (d) with $\gamma=\frac{1-\beta}{2}\ge 0$, we obtain (2.45) with $\sigma_2=\sigma$ given in (2.19) in Theorem 2.3.3.

For (2.46), it follows from the expression of $q_j(\varepsilon)$ in (2.32) that

$$\frac{|\,\mathbf{W}_j^{R(\varepsilon)}\,|}{q_j(\varepsilon)}=\frac{M^{\frac{\beta+1}{2}(j-1)}}{\theta\mathbf{h}^{\frac{\beta}{2}}\underline{C}_{M,\beta}\mu^*(\varepsilon)}.$$

Owing to the definition of Z_j in (2.14) and to inequality (2.33), we get

$$\frac{1}{\varepsilon^2}\sum_{j=2}^{R(\varepsilon)}\frac{\left(\mathbf{W}_j^{R(\varepsilon)}\right)^2}{\left(q_j(\varepsilon)N(\varepsilon)\right)^2}\operatorname{Var}\left(\widetilde{Y}_j\right) \le \frac{\mathbf{h}^\beta}{\left(\theta\mathbf{h}^{\frac{\beta}{2}}\underline{C}_{M,\beta}\underline{\mu}^*\right)^2}\frac{1}{\left(\varepsilon N(\varepsilon)\right)^2}\sum_{j=2}^{R(\varepsilon)}M^{(j-1)}\operatorname{Var}(Z_j)$$
$$\le \frac{\mathbf{h}^\beta}{\left(\theta\mathbf{h}^{\frac{\beta}{2}}\underline{C}_{M,\beta}\underline{\mu}^*\right)^2}\frac{1}{\left(\varepsilon N(\varepsilon)\right)^2}\left(\sup_{j\ge1}\operatorname{Var}(Z_j)\right)\frac{M^{R(\varepsilon)-1}}{M-1}.$$

We conclude by showing that

$$\frac{M^{R(\varepsilon)}}{\left(\varepsilon N(\varepsilon)\right)^2}\to 0, \quad \text{as} \quad \varepsilon\to 0. \tag{2.47}$$

Owing to the expression of $R(\varepsilon)$ given in (2.10), we notice that

$$R(\varepsilon)=O\left(\sqrt{\log(1/\varepsilon)}\right)=o\Big(\log(1/\varepsilon)\Big) \quad \text{as} \quad \varepsilon\to 0.$$

Moreover, using Lemma 2.4.3, up to another reduction of $\bar{\varepsilon}$, we have $\frac{1}{(\varepsilon N(\varepsilon))^2}\le\left(\frac{2}{C_\beta}\right)^2\varepsilon^2$ for all $\beta>0$. This in turn yields

$$\frac{M^{R(\varepsilon)}}{\left(\varepsilon N(\varepsilon)\right)^2}\le\left(\frac{2}{C_\beta}\right)^2\varepsilon^2M^{R(\varepsilon)}=C_\beta^{-2}e^{\log(M)R(\varepsilon)-2\log(1/\varepsilon)}\to 0, \quad \text{as} \quad \varepsilon\to 0.$$

Then (2.46) is proved and so is the first condition of Lindeberg's Theorem.

For the second condition of Lindeberg's Theorem we need to prove that, for every $\eta>0$,

$$\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\mathbf{E}\left[\left(\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right)^2\mathbf{1}_{\left\{\left|\frac{1}{\varepsilon}\frac{\mathbf{w}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right|>\eta\right\}}\right]\xrightarrow{\varepsilon\to0}0. \tag{2.48}$$

Since the $\left(\widetilde{Y}_j^k\right)_{k=1,\dots,N_j(\varepsilon)}$ are identically distributed, we can write

$$\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\mathbf{E}\left[\left(\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right)^2\mathbf{1}_{\left\{\left|\frac{1}{\varepsilon}\frac{\mathbf{w}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right|>\eta\right\}}\right]$$
$$\le\sum_{j=2}^{R(\varepsilon)}\frac{1}{\varepsilon^2}\frac{\left|\mathbf{W}_j^{R(\varepsilon)}\right|^2}{N_j(\varepsilon)}\mathbf{E}\left[\left(\widetilde{Y}_j\right)^2\mathbf{1}_{\left\{|\widetilde{Y}_j|>\eta\,\varepsilon\frac{N_j(\varepsilon)}{\left|\mathbf{w}_j^{R(\varepsilon)}\right|}\right\}}\right].$$

We set $\widetilde{Z}_j=Z_j-\mathbf{E}\left[Z_j\right]$. Replacing q_j by its values given in (2.32), using Inequality (2.30) from Lemma 2.4.1 (b) and the elementary inequality $\frac{1}{N_j(\varepsilon)}\le\frac{1}{q_j(\varepsilon)N(\varepsilon)}$, yields

$$\sum_{j=2}^{R(\varepsilon)}\sum_{k=1}^{N_j(\varepsilon)}\mathbf{E}\left[\left(\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right)^2\mathbf{1}_{\left\{\left|\frac{1}{\varepsilon}\frac{\mathbf{w}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\right|>\eta\right\}}\right]$$

$$\leq \frac{1}{\theta\mathbf{h}^{\frac{\beta}{2}}\underline{C}_{M,\beta}\mu^*(\varepsilon)}\frac{1}{\varepsilon^2N(\varepsilon)}\sum_{j=2}^{R(\varepsilon)}\left|\mathbf{W}_j^{R(\varepsilon)}\right|M^{\frac{\beta+1}{2}(j-1)}\mathbf{E}\left[\left(\widetilde{Y}_j\right)^2\mathbf{1}_{\left\{|\widetilde{Y}_j|>\eta\frac{\theta\mathbf{h}^{\frac{\beta}{2}}\underline{C}_{M,\beta}\mu^*(\varepsilon)}{M^{\frac{\beta+1}{2}}(j-1)}\varepsilon N(\varepsilon)\right\}}\right]$$

$$\leq \frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta\underline{C}_{M,\beta}\underline{\mu}^*}a_\infty\widetilde{B}_\infty\frac{1}{\varepsilon^2N(\varepsilon)}\sum_{j=2}^{R(\varepsilon)}M^{\frac{1-\beta}{2}(j-1)}\mathbf{E}\left[\left(\widetilde{Z}_j\right)^2\mathbf{1}_{\left\{|\widetilde{Z}_j|>\eta\theta\underline{C}_{M,\beta}\underline{\mu}^*\varepsilon N(\varepsilon)M^{-\frac{j-1}{2}}\right\}}\right]$$

$$\leq \frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta\underline{C}_{M,\beta}\underline{\mu}^*}a_\infty\widetilde{B}_\infty\sup_{2\leq j\leq R(\varepsilon)}\mathbf{E}\left[(Z_j)^2\mathbf{1}_{\left\{|\widetilde{Z}_j|>\Theta\varepsilon N(\varepsilon)M^{-\frac{R(\varepsilon)}{2}}\right\}}\right]\frac{1}{\varepsilon^2N(\varepsilon)}\sum_{j=2}^{R(\varepsilon)}M^{\frac{1-\beta}{2}(j-1)}$$

where we set $\Theta=\eta\theta\underline{C}_{M,\beta}\underline{\mu}^*\sqrt{M}$. Now, it follows from Lemma 2.4.3 that

$$\frac{1}{\varepsilon^2N(\varepsilon)}\sum_{j=2}^{R(\varepsilon)}M^{\frac{1-\beta}{2}(j-1)}$$

$$=\frac{1}{\varepsilon^2N(\varepsilon)}\left(\frac{M^{\frac{1-\beta}{2}R(\varepsilon)}-M^{\frac{1-\beta}{2}}}{M^{\frac{1-\beta}{2}}-1}\mathbf{1}_{\{\beta\neq1\}}+(R(\varepsilon)+1)\mathbf{1}_{\{\beta=1\}}\right)\to K,$$

as $\varepsilon\to0$, where K is a real positive constant. Owing to (2.47) $\lim_{\varepsilon\to0}\varepsilon N(\varepsilon)M^{-\frac{R(\varepsilon)}{2}}=+\infty$. Hence, since we assumed that the family $(Z_j)_{j\geq1}$ is L^2–uniformly integrable, we obtain that

$$\lim_{\varepsilon\to0}\sup_{2\leq j\leq R(\varepsilon)}\mathbf{E}\left[(Z_j)^2\mathbf{1}_{\left\{|Z_j|>\Theta\varepsilon N(\varepsilon)M^{-\frac{R(\varepsilon)-1}{2}}\right\}}\right]=0 \tag{2.49}$$

and the second condition of Lindeberg's Theorem is proved.

▷ MLMC : The proofs are quite the same as for ML2R, up to the constant $1+\frac{1}{2\alpha}$, coming from the constant C_β in the asymptotic of $N(\varepsilon)$. Using Lemma 2.4.3 and the expression of $R(\varepsilon)$ given in (2.11), we obtain

$$N(\varepsilon)\overset{\varepsilon\to0}{\sim}C_\beta\varepsilon^{-2}\begin{cases}1 & \text{if } \beta>1,\\ \frac{1}{\alpha\log(M)}\log\left(\frac{1}{\varepsilon}\right) & \text{if } \beta=1,\\ \varepsilon^{-\frac{1-\beta}{2\alpha}} & \text{if } \beta<1.\end{cases}\qquad(MLMC)$$

We replace $\mathbf{W}_j^R=1$, $j=1,\dots,R$ and $a_\infty\widetilde{B}_\infty=1$. The only significant difference comes when $\beta<1$, while proving (2.47). In this case, owing to Lemma 2.4.3 as we did in (2.41) and using the expression of $R(\varepsilon)$ given in (2.11), up to reducing $\bar{\varepsilon}$, we can write

$$\frac{M^{R(\varepsilon)}}{(\varepsilon N(\varepsilon))^2}\leq\left(\frac{2}{C_\beta}\right)^2\varepsilon^2M^{-(1-\beta)R(\varepsilon)}M^{R(\varepsilon)}\leq\left(\frac{2}{C_\beta}\right)^2M^{\beta(C_R^{(1)}+1)}\varepsilon^{2-\frac{\beta}{\alpha}}$$

which goes to 0, owing to the strict inequality assumption $2\alpha>\beta$. □

2.6 Applications

2.6.1 Diffusions

In this section we retrieve a recent result by Kebaier and Ben Alaya (see [BAK15]) obtained for MLMC estimators and we extend it to the ML2R estimators and to the use of path-dependent functionals. Let $(X_t)_{t\in[0,T]}$ a Brownian diffusion process solution to the stochastic differential equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)\mathrm{d}W_s, \quad t \in [0,T]$$

where $b : [0,T] \times \mathbf{R}^d \to \mathbf{R}^d$, $\sigma : [0,T] \times \mathbf{R}^d \to M(d,q,\mathbf{R})$ are continuous functions, Lipschitz continuous in x, uniformly in $t \in [0,T]$, $(W_t)_{t\in[0,T]}$ is a q-dimensional Brownian motion independent of X_0, both defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$.

We know that $X = (X_t)_{t\in[0,T]}$ is the unique $(\mathcal{F}_t^W)_{t\in[0,T]}$-adapted solution to this equation, where $\mathcal{F}^W$ is the augmented filtration of W. The process $(X_t)_{t\in[0,T]}$ cannot be simulated at a reasonable computational cost (at least in full generality), which leads to introduce some simulatable time discretization schemes, the simplest being undoubtedly the Euler scheme with step $h = \frac{T}{n}$, $n \geq 1$, defined by

$$\bar{X}_t^n = X_0 + \int_0^t b(\underline{s}, \bar{X}_{\underline{s}}^n)ds + \int_0^t \sigma(\underline{s}, \bar{X}_{\underline{s}}^n)\mathrm{d}W_s \tag{2.50}$$

with $\underline{s} = \lfloor ns/T \rfloor \frac{T}{n}$, $s \in [0,T]$. In particular, if we set $t_k^n = k\frac{T}{n}$,

$$\bar{X}_{t_{k+1}^n}^n = \bar{X}_{t_k^n}^n + b(t_k^n, \bar{X}_{t_k^n}^n)h + \sigma(t_k^n, \bar{X}_{t_k^n}^n)\sqrt{h}U_{k+1}^n, \quad k \in \{0, \ldots, n-1\}$$

where $U_{k+1}^n = \frac{W_{t_{k+1}^n} - W_{t_k^n}}{\sqrt{h}}$ is *i.i.d.* with distribution $\mathcal{N}(0, I_q)$. Furthermore, we also derive from (2.50) that

$$\bar{X}_t^n = \bar{X}_{\underline{t}}^n + b(\underline{t}, \bar{X}_{\underline{t}}^n)(t - \underline{t}) + \sigma(\underline{t}, \bar{X}_{\underline{t}}^n)(W_t - W_{\underline{t}}), \quad t \in [0,T].$$

It is classical background that, under the above assumptions on b, σ, X_0 and W, the Euler scheme satisfies the following a priori L^p-error bounds :

$$\forall p \geqslant 2,\ \exists c_{b,\sigma,p,T} > 0, \quad \left\| \sup_{t\in[0,T]} |X_t - \bar{X}_t^n| \right\|_p \leq c_{b,\sigma,p,T}\sqrt{\frac{T}{n}}\left(1 + \left\|X_0\right\|_p\right). \tag{2.51}$$

For the weak error expansion the existing results are less general. Let us recall as an illustration the celebrated Talay-Tubaro's and Bally-Talay's weak error expansions for marginal functionals of Brownian diffusions, i.e. functionals of the form $F(X) = f(X_T)$.

Theorem 2.6.1. *The following statements hold. (a) Regular setting (Talay-Tubaro [TT90]) : If b and σ are infinitely differentiable with bounded partial derivatives and if $f : \mathbf{R}^d \to \mathbf{R}$ is an infinitely differentiable function, with all its partial derivatives having a polynomial growth, then for a fixed maturity $T > 0$ and for every integer $R \in N^*$*

$$\mathbf{E}\left[f(\bar{X}_T^n)\right] - \mathbf{E}\left[f(X_T)\right] = \sum_{k=1}^{R} c_k \left(\frac{1}{n}\right)^k + O\left(\left(\frac{1}{n}\right)^{R+1}\right) \tag{2.52}$$

where the coefficients c_k depend on b, σ, f, T but not on n.
(b) (Hypo-)Elliptic setting (Bally-Talay [BT96a]) : If b and σ are infinitely differentiable with bounded partial derivatives and if σ is uniformly elliptic in the sense that

$$\forall x \in \mathbf{R}^d,\ \forall t \in [0,T], \quad \sigma\sigma^*(x) \geq \varepsilon_0 I_q,\ \varepsilon_0 > 0$$

or more generally if (b,σ) satisfies the strong Hörmander hypo-ellipticity assumption, then (2.52) holds true for every bounded Borel function $f : \mathbf{R}^d \to \mathbf{R}$.

For more general path-dependent functionals, no such result exists in general. For various classes of specified functionals depending on the running maximum or mean, some exit stopping time, first order weak expansions in $h^\alpha, \alpha \in (0,1]$, have sometimes been established (see [LP17] for a brief review in connection with multilevel methods). However, as emphasized by the numerical experiments carried out in [LP17], such weak error expansion can be highly suspected to hold at any order under reasonable smoothness assumptions.

In this section we consider $F : \mathcal{C}_b([0,T],\mathbf{R}^d) \to \mathbf{R}$ a Lipschitz continuous functional and we set

$$Y_0 = F(X) \quad \text{and} \quad Y_h = F(\bar{X}^n) \text{ with } h = \frac{T}{n} \text{ and } n \geqslant 1 \ (i.e.\ \mathbf{h} = T).$$

We assume the weak error expansion ($WE_{\alpha,\bar{R}}$). We prove now that both estimators ML2R (2.3) and MLMC (2.2) satisfy a Strong Law of Large Numbers and a Central Limit Theorem when ε tends to 0.

Theorem 2.6.2. *Let $X_0 \in L^2$ and assume that $F : \mathcal{C}_b([0,T],\mathbf{R}^d) \to \mathbf{R}$ is a Lipschitz continuous functional. Then the assumption (SE_β) is satisfied with $\beta = 1$.*

If $X_0 \in L^p$ for $p \geqslant 2$, then the L^p–strong error assumption $\|Y_h - Y_0\|_p \leqslant V_1^{(p)}\sqrt{h}$ is satisfied so that both ML2R and MLMC estimators satisfy Theorem 2.3.1.

If $X_0 \in L^p$ for $p > 2$ and if F is differentiable with DF continuous, then the sequence $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable and

$$\exists v_\infty > 0, \quad \lim_{h\to 0} \|Z(h)\|_2^2 = (M-1)v_\infty. \tag{2.53}$$

As a consequence, both ML2R and MLMC estimators satisfy Theorem 2.3.3 (case $\beta = 1$).

Proof. First, note that if F is a Lipschitz continuous functional, with Lipschitz coefficient $[F]_{\mathrm{Lip}}$, we have for all $p \geqslant 2$

$$\|Y_h - Y_0\|_p^p \leqslant [F]_{\mathrm{Lip}}^p \mathbf{E}\left[\sup_{t\in[0,T]} |X_t - \bar{X}_t^n|^p\right] \leqslant [F]_{\mathrm{Lip}}^p c_{b,\sigma,p,T}^p (1 + \|X_0\|_p)^p h^{\frac{p}{2}},$$

then $(Y_h)_{h\in\mathcal{H}}$ satisfies (SE_β) with $\beta = 1$ and the L^p–strong error assumption as soon as $X_0 \in L^p$.

Assume now that $X_0 \in L^p$ for $p > 2$. By a straightforward application of Minkowski's inequality we deduce from the L^p–strong error assumption that $\|Y_{\frac{h}{M}} - Y_h\|_p \leqslant C\sqrt{h}$ and then that $\sup_{h\in\mathcal{H}} \|Z(h)\|_p < +\infty$. Applying the criterion (a) of Lemma 2.5.1 we prove that $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable.

At this stage it remains to prove (2.53). The key is Theorem 3 in [BAK15], where it is proved that

$$\sqrt{nM}\left(\bar{X}^n - \bar{X}^{nM}\right) \xrightarrow{stably} U^{(M)}, \quad \text{as } n \to +\infty,$$

where $U^{(M)} = (U_t^{(M)})_{t\in[0,T]}$ is the $d-$dimensional process satisfying

$$U_t^{(M)} = \sqrt{\frac{M-1}{2}} \sum_{i,j=1}^{q} V_t \int_0^t (V_s)^{-1} \nabla\varphi_{.j}(X_s)\varphi_{.i}(X_s) \mathrm{d}B_s^{i,j}, \quad t \in [0,T]. \tag{2.54}$$

We recall the notations of Jacod and Protter [JP98]

$$dX_t = \varphi(X_t)dW_t = \sum_{j=0}^{q} \varphi_{.j}(X_t)\mathrm{d}W_t^j$$

with $\varphi_{.j}$ representing the jth column of the matrix $\varphi = [\varphi_{ij}]_{\substack{i=1,\dots,d,\\ j=1,\dots,q}}$, for $j = 1,\dots,q$, $\varphi_0 = b$ and $W_t := (t, W_t^1, \dots, W_t^q)'$ (column vector), where $W_t^0 = t$ and the q remaining components make up a standard Brownian motion. Moreover, $\nabla\varphi_{.j}$ is a $d \times d$ matrix where $(\nabla\varphi_{.j})_{ik} = \partial_{x^k}\varphi_{ij}$ (partial derivative of φ_{ij} with respect to the kth coordinate) and $(V_t)_{t\in[0,T]}$ is the $\mathbf{R}^{d\times d}$ valued process solution of the linear equation

$$V_t = I_d + \sum_{j=0}^{q} \int_0^t \nabla\varphi_{.j}(X_s)V_s \mathrm{d}W_s^j, \quad t \in [0,T].$$

Here $(B^{ij})_{1\le i,j\le q}$ is a standard q^2-dimensional Brownian motion independent of W. This process is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}_t})_{t\ge0}, \widetilde{\mathbf{P}})$ of the original space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbf{P})$ on which lives W.

We write, using that $h = \frac{T}{n}$,

$$Z(h) = \sqrt{nM}\left(F(\bar{X}^{nM}) - F(\bar{X}^n)\right) = -\int_0^1 DF\left(u\bar{X}^n + (1-u)\bar{X}^{nM}\right)\mathrm{d}u \cdot U_n^{(M)}$$

where $U_n^{(M)} := \sqrt{nM}\left(\bar{X}^n - \bar{X}^{nM}\right)$. The function $(x_1,x_2,x_3) \mapsto \int_0^1 DF(ux_1 + (1-u)x_2)\mathrm{d}ux_3$ is continuous, and it suffices to prove that $\left(\bar{X}^n, \bar{X}^{nM}, U_n^{(M)}\right) \xrightarrow{\mathcal{L}} (X, X, U^{(M)})$, as n goes to infinity, to conclude that

$$Z(h) \xrightarrow{\mathcal{L}} -DF(X)U^{(M)}, \quad \text{as } h \to 0. \tag{2.55}$$

Let two bounded Lipschitz continuous functionals be $\phi : \mathcal{C}_b([0,T], \mathbf{R}^{2d}) \to \mathbf{R}$ and $\psi : \mathcal{C}_b([0,T], \mathbf{R}^d) \to \mathbf{R}$ and let denote $\widetilde{X}^n = (\bar{X}^n, \bar{X}^{nM})$ and $\widetilde{X} = (X, X)$. We write

$$\mathbf{E}\left[\phi(\widetilde{X}^n)\psi(U_n^{(M)}) - \phi(\widetilde{X})\psi(U^{(M)})\right] = \mathbf{E}\left[(\phi(\widetilde{X}^n) - \phi(\widetilde{X}))\psi(U_n^{(M)}) + \phi(\widetilde{X})(\psi(U_n^{(M)}) - \psi(U^{(M)}))\right]$$

Since $(U_n^{(M)})_{n\geqslant1}$ converges stably with limit $U^{(M)}$, we have that

$$\lim_{n\to+\infty} \mathbf{E}\left[\phi(\widetilde{X})(\psi(U_n^{(M)}) - \psi(U^{(M)}))\right] = 0.$$

On the other hand, owing to (2.51), we prove that

$$\lim_{n\to+\infty} \mathbf{E}\left[(\phi(\widetilde{X}^n) - \phi(\widetilde{X}))\psi(U_n^{(M)})\right] = 0.$$

By (2.55) and Lemma 2.5.1 (b) we have

$$\lim_{h\to0} \left\|Z(h)\right\|_2^2 = \left\|DF(X)U^{(M)}\right\|_2^2 = (M-1)v_\infty$$

with $v_\infty = \left\|DF(X)\frac{U^{(M)}}{M-1}\right\|_2^2$ which does not depend on M owing to the definition of U^M given in (2.54). □

2.6.2 Nested Monte Carlo

The aim of a *nested* Monte Carlo method is to compute by Monte Carlo simulation

$$\mathbf{E}\left[f(\mathbf{E}\left[X \mid Y\right])\right]$$

where (X, Y) is a couple of $\mathbf{R} \times \mathbf{R}^{q_Y}$-valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with $X \in L^2(\mathbf{P})$ and $f : \mathbf{R} \to \mathbf{R}$ is a Lipschitz continuous function with Lipschitz coefficient $[f]_{\text{Lip}}$. We assume that there exists a Borel function $F : \mathbf{R}^{q_\xi} \times \mathbf{R}^{q_Y} \to \mathbf{R}$ and a random variable $\xi : (\Omega, \mathcal{A}) \to \mathbf{R}^{q_\xi}$ independent of Y such that

$$X = F(\xi, Y)$$

and we set $\mathbf{h} = \frac{1}{K_0}$ for some integer $K_0 \geqslant 1$, $h = 1/K$, $K \in K_0\mathbf{N}^* = \left\{K_0, 2K_0, \ldots\right\}$ and

$$Y_0 := f(\mathbf{E}\left[X \mid Y\right]), \quad Y_h = Y_{\frac{1}{K}} := f\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi_k, Y)\right) \tag{2.56}$$

where $(\xi_k)_{k\geq 1}$ is a sequence of *i.i.d.* variables, $\xi_k \sim \xi$, independent of Y. A *nested* ML2R estimator then writes ($n_j = M^{j-1}$)

$$\begin{aligned} I_\pi^N &= \frac{1}{N_1}\sum_{i=1}^{N_1} f\left(\frac{1}{K}\sum_{k=1}^{K} F\left(\xi_k^{(1),i}, Y^{(1),i}\right)\right) \\ &+ \sum_{j=2}^{R} \frac{\mathbf{W}_j^R}{N_j}\sum_{i=1}^{N_j}\left(f\left(\frac{1}{n_j K}\sum_{k=1}^{n_j K} F\left(\xi_k^{(j),i}, Y^{(j),i}\right)\right) - f\left(\frac{1}{n_{j-1}K}\sum_{k=1}^{n_{j-1}K} F\left(\xi_k^{(j),i}, Y^{(j),i}\right)\right)\right) \end{aligned} \tag{2.57}$$

where $\left(Y^{(j),i}\right)_{i\geq 1}$ is a sequence of independent copies of $Y^{(j)} \sim Y$, $j = 1, \ldots, R$, $Y^{(j)}$ independent of $Y^{(\ell)}$ for $j \neq \ell$, and $\left(\xi_k^{(j),i}\right)_{k,i\geq 1, j=1,\ldots,R}$ is a sequence of *i.i.d.* variables $\xi_k^{(j),i} \sim \xi$. We saw in [LP17] that, when f is $2R$ times differentiable with $f^{(k)}$ bounded, the *nested* Monte Carlo estimator satisfies (SE_β) with $\beta = 1$ and ($WE_{\alpha,\bar{R}}$) with $\alpha = 1$ and $\bar{R} = R - 1$. Here we want to show that the *nested* Monte Carlo satisfies also the assumptions of the Strong Law of Large Numbers 2.3.1 and of the Central Limit Theorem 2.3.3. Then, we define for convenience

$$\phi_0(y) := \mathbf{E}\left[F(\xi, y)\right], \quad \phi_h(y) := \frac{1}{K}\sum_{k=1}^{K} F(\xi_k, y), \quad K \in K_0\mathbf{N}^*, \tag{2.58}$$

so that $Y_0 = f(\phi_0(Y))$ and $Y_h = f(\phi_h(Y))$, and for a fixed y, we set $\sigma_F(y) := \sqrt{\text{Var}(F(\xi, y))}$.

Proposition 2.6.1. *Still assuming that f is Lispchitz continuous. If $X \in L^p(\mathbf{P})$ for $p \geqslant 2$, then there exists $V_1^{(p)}$ such that, for all $h = \frac{1}{K}$ and $h' = \frac{1}{K'}, K, K' \in K_0\mathbf{N}^*$,*

$$\left\|Y_{h'} - Y_h\right\|_p^p \leq V_1^{(p)}\left|h' - h\right|^{\frac{p}{2}}. \tag{2.59}$$

As a consequence, the assumption (SE_β) and the L^p–strong error assumption (2.12) are satisfied with $\beta = 1$. Then both ML2R and MLMC estimators satisfy a Strong Law of Large Numbers, see Theorem 2.3.1.

Proof. Set $\widetilde{X}_k = F(\xi_k, Y) - \mathbf{E}\left[F(\xi_k, Y)\,|\,Y\right]$ and $S_k = \sum_{\ell=1}^{k} \widetilde{X}_\ell$. As f is Lipschitz, we have

$$\begin{aligned}\left\|Y_{h'} - Y_h\right\|_p^p &= \left\| f\left(\frac{1}{K'}\sum_{k=1}^{K'} F(\xi_k, Y)\right) - f\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi_k, Y)\right)\right\|_p^p \\ &\le [f]_{\mathrm{Lip}}^p \left\| \frac{1}{K'}\sum_{k=1}^{K'} \widetilde{X}_k - \frac{1}{K}\sum_{k=1}^{K} \widetilde{X}_k \right\|_p^p = [f]_{\mathrm{Lip}}^p\, \mathbf{E}\left[\left|\frac{S_{K'}}{K'} - \frac{S_K}{K}\right|^p\right].\end{aligned}$$

Assume without loss of generality that $K \le K'$. Since $p \ge 2$, it follows that

$$\begin{aligned}\mathbf{E}\left[\left|\frac{S_{K'}}{K'} - \frac{S_K}{K}\right|^p\right] &= \mathbf{E}\left[\left|\left(\frac{1}{K'} - \frac{1}{K}\right) S_K + \frac{1}{K'}(S_{K'} - S_K)\right|^p\right] \\ &\le 2^{p-1}\left[\left|\frac{1}{K'} - \frac{1}{K}\right|^p \mathbf{E}\left[|S_K|^p\right] + \left(\frac{1}{K'}\right)^p \mathbf{E}\left[|S_{K'} - S_K|^p\right]\right].\end{aligned}$$

Owing to Burkholder's inequality, there exists a universal constant C_p such that

$$\mathbf{E}\left[|S_K|^p\right] \le C_p \mathbf{E}\left[\left|\sum_{k=1}^{K} \widetilde{X}_k^2\right|^{\frac{p}{2}}\right] \le C_p \left(\sum_{k=1}^{K} \left\|\widetilde{X}_k^2\right\|_{\frac{p}{2}}\right)^{\frac{p}{2}} = C_p K^{\frac{p}{2}} \mathbf{E}\left[\left|\widetilde{X}_1\right|^p\right].$$

Hence, as $S_{K'} - S_K \sim S_{K'-K}$ in distribution,

$$\mathbf{E}\left[\left|\frac{S_{K'}}{K'} - \frac{S_K}{K}\right|^p\right] \le 2^{p-1} C_p \mathbf{E}\left[|\widetilde{X}_1|^p\right]\left[\left|\frac{1}{K'} - \frac{1}{K}\right|^p K^{\frac{p}{2}} + \left(\frac{1}{K'}\right)^p |K' - K|^{\frac{p}{2}}\right].$$

Keeping in mind that $K' \ge K$, we derive

$$\begin{aligned}\left|\frac{1}{K'} - \frac{1}{K}\right|^p K^{\frac{p}{2}} + \left|\frac{1}{K'}\right|^p |K' - K|^{\frac{p}{2}} &= \left|\frac{1}{K'} - \frac{1}{K}\right|^{\frac{p}{2}} \left|\frac{K}{K'} - 1\right|^{\frac{p}{2}} + \left|\frac{K}{K'}\right|^{\frac{p}{2}} \left|\frac{1}{K} - \frac{1}{K'}\right|^{\frac{p}{2}} \\ &\le 2\left|\frac{1}{K} - \frac{1}{K'}\right|^{\frac{p}{2}}.\end{aligned}$$

We conclude by setting $V_1^{(p)} = [f]_{\mathrm{Lip}}^p\, 2^p C_p \mathbf{E}\left[|\widetilde{X}_1|^p\right]$. □

For the Central Limit Theorem to hold, the key point is the following Lemma.

Lemma 2.6.3. *Assume that $f : \mathbf{R} \to \mathbf{R}$ is a Lipschitz continuous function and differentiable with f' continuous. Let ζ be an $\mathcal{N}(0,1)$-distributed random variable independent of Y. Then, as $h \to 0$,*

$$Z(h) = \sqrt{\frac{M}{h}}\left(Y_{\frac{h}{M}} - Y_h\right) \xrightarrow{\mathcal{L}} \sqrt{M-1} f'(\phi_0(Y))\sigma_F(Y)\zeta. \tag{2.60}$$

Proof. First note that $Z(h) = z_h^{(M)}(Y)$ where $z_h^{(M)}$ is defined by

$$\forall y \in \mathbf{R}^{q_Y}, \quad z_h^{(M)}(y) = \sqrt{\frac{M}{h}}\left(f(\phi_{\frac{h}{M}}(y)) - f(\phi_h(y))\right).$$

Let $y \in \mathbf{R}^{q_Y}$. We have

$$z_h^{(M)}(y) = -\left(\int_0^1 f'\left(v\phi_h(y) + (1-v)\phi_{\frac{h}{M}}(y)\right) \mathrm{d}v\right) u_h^{(M)}(y) \tag{2.61}$$

with

$$u_h^{(M)}(y) = \sqrt{\frac{M}{h}}\left(\phi_h(y) - \phi_{\frac{h}{M}}(y)\right).$$

We derive from the Strong Law of Large Numbers that

$$\lim_{h\to 0}\phi_h(y) = \phi_0(y) = \lim_{h\to 0}\phi_{\frac{h}{M}}(y) \quad \text{a.s.}$$

and by continuity of the function $(x_1, x_2) \mapsto \int_0^1 f'(vx_1+(1-v)x_2)\mathrm{d}v$ (since f' is continuous) we get

$$\lim_{h\to 0}\int_0^1 f'\left(v\phi_h(y) + (1-v)\phi_{\frac{h}{M}}(y)\right)\mathrm{d}v = f'\left(\phi_0(y)\right) \quad a.s. \tag{2.62}$$

We have now to study the convergence of the random sequence $u_h^{(M)}(y)$ as h goes to zero. We set $\widetilde{\xi}_k = \xi_{k+K}$, $k = 1, \ldots, K(M-1)$. Note that $(\widetilde{\xi}_k)_{k=1,\ldots,K(M-1)}$ are *i.i.d.* with distribution ξ_1 and are independent of $(\xi_k)_{k=1\ldots,M}$. Then we can write

$$\begin{aligned}
u_h^{(M)}(y) &= \sqrt{MK}\left(\frac{1}{K}\sum_{k=1}^{K}F(\xi_k, y) - \frac{1}{MK}\sum_{k=1}^{MK}F(\xi_k, y)\right)\\
&= \sqrt{MK}\left(\frac{M-1}{MK}\sum_{k=1}^{K}\left(F(\xi_k, y) - \phi_0(y)\right) - \frac{1}{MK}\sum_{k=K+1}^{MK}\left(F(\xi_k, y) - \phi_0(y)\right)\right)\\
&= \frac{M-1}{\sqrt{M}}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}F(\xi_k, y) - \phi_0(y)\right)\\
&\qquad - \sqrt{\frac{M-1}{M}}\left[\frac{1}{\sqrt{K(M-1)}}\left(\sum_{k=1}^{K(M-1)}F(\widetilde{\xi}_k, y) - \phi_0(y)\right)\right].
\end{aligned}$$

Owing to the Central Limit Theorem and the independence of both terms in the right hand side of the above inequality, we derive that

$$u_h^{(M)}(y) \xrightarrow{\mathcal{L}} \frac{M-1}{\sqrt{M}}\sigma_F(y)\zeta_1 - \sqrt{\frac{M-1}{M}}\sigma_F(y)\zeta_2, \quad \text{as } h \to 0,$$

where ζ_1 and ζ_2 are two independent random variables both following a standard Gaussian distribution. Hence, noting that $\left(\frac{M-1}{\sqrt{M}}\right)^2 + \left(\sqrt{\frac{M-1}{M}}\right)^2 = M-1$, we obtain

$$u_h^{(M)}(y) \xrightarrow{\mathcal{L}} \sqrt{M-1}\sigma_F(y)\zeta \quad \text{with} \quad \zeta \sim \mathcal{N}(0,1). \tag{2.63}$$

By Slutsky's Theorem, we derive from (2.61), (2.62) and (2.63) that for every $y \in \mathbf{R}^{q_Y}$,

$$z_h^{(M)}(y) \xrightarrow{\mathcal{L}} \sqrt{M-1}f'\left(\phi_0(y)\right)\sigma_F(y)\zeta, \quad \text{as } h \to 0. \tag{2.64}$$

Recall that $Z(h) = z_h^{(M)}(Y)$. We prove (2.60) combining Fubini's theorem with Lebesgue dominated convergence theorem and (2.64). More precisely, for all $G \in \mathcal{C}_b$ we have

$$\begin{aligned}
\lim_{h\to 0}\mathbf{E}\Big[G(Z(h))\Big] &= \lim_{h\to 0}\mathbf{E}\left[G\left(z_h^{(M)}(Y)\right)\right] = \mathbf{E}\left[\lim_{h\to 0}G\left(z_h^{(M)}(Y)\right)\right]\\
&= \mathbf{E}\left[G\left(\sqrt{M-1}f'\left(\phi_0(y)\right)\sigma_F(y)\zeta\right)\right],
\end{aligned}$$

as desired. □

We are now in a position to prove that the *nested* Monte Carlo satisfies the assumptions of the Central Limit Theorem 2.3.3.

Theorem 2.6.4. *Assume that $f : \mathbf{R} \to \mathbf{R}$ is a Lipschitz continuous function and differentiable with f' continuous. Then $\big(Z(h)\big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable and*

$$\lim_{h\to 0} \big\|Z(h)\big\|_2^2 = (M-1)\big\|f'(\phi_0(Y))\sigma_F(Y)\big\|_2^2 \tag{2.65}$$

As a consequence, the ML2R and MLMC estimators (2.3) and (2.2) satisfy a Central Limit Theorem in the sense of Theorem 2.3.3 (case $\beta = 1$).

Proof. We prove first the L^2–uniform integrability of $\big(Z(h)\big)_{h\in\mathcal{H}}$. As f is Lipschitz we have,

$$\big|Z(h)\big|^2 \leqslant [f]_{\mathrm{Lip}}^2 \big|u_h^{(M)}(Y)\big|^2, \quad \text{with} \quad u_h^{(M)}(y) = \sqrt{\frac{M}{h}}\left(\phi_h(y) - \phi_{\frac{h}{M}}(y)\right).$$

Consequently it suffices to show that $\Big(u_h^{(M)}(Y)\Big)_{h\in\mathcal{H}}$ is L^2–uniformly integrable, to establish the L^2–uniform integrability of $\big(Z(h)\big)_{h\in\mathcal{H}}$.

We saw in the proof of Proposition 2.6.3 that $u_h^{(M)}(Y) \xrightarrow{\mathcal{L}} \sqrt{M-1}\sigma_F(Y)\zeta$ as h goes to 0, where ζ is a standard normal random variable independent of Y. Owing to Lemma 2.5.1 (b), the uniform integrability will follow from $\lim_{h\to 0} \big\|u_h^{(M)}(Y)\big\|_2 = \big\|\sqrt{M-1}\sigma_F(Y)\zeta\big\|_2$. In fact this convergence holds as an equality. Indeed

$$\begin{aligned}\big\|u_h^{(M)}(Y)\big\|_2^2 &= MK\mathbf{E}\left[(S_K - S_{MK})^2\right] \\ &= MK\mathbf{E}\left[\left(\frac{M-1}{MK}S_K - \frac{1}{MK}(S_{MK} - S_K)\right)^2\right].\end{aligned}$$

We notice that $S_{MK} - S_K$ is independent of S_K. Hence, since the ξ_k are independent,

$$\begin{aligned}\big\|u_h^{(M)}(Y)\big\|_2^2 &= MK\left(\mathbf{E}\left[\left(\frac{M-1}{MK}S_K\right)^2\right] + \mathbf{E}\left[\left(\frac{1}{MK}(S_{MK} - S_K)\right)^2\right]\right) \\ &= MK\left(\left(\frac{M-1}{MK}\right)^2 K\mathbf{E}\left[(\widetilde{X}_1)^2\right] + \left(\frac{1}{MK}\right)^2 (MK-K)\mathbf{E}\left[\left(\widetilde{X}_1\right)^2\right]\right) \\ &= (M-1)\mathbf{E}\left[(\widetilde{X}_1)^2\right] = (M-1)\mathbf{E}\left[\sigma_F^2(Y)\right].\end{aligned}$$

We prove now (2.65) using again the Lemma 2.5.1 (b) with the convergence in law of $\big(Z(h)\big)_{h\in\mathcal{H}}$ established in Lemma 2.6.3. □

We notice that, if the assumption (2.59) in Proposition 2.6.1 holds with $p > 2$, the condition of L^2–uniform integrability is much easier to show since it is a direct consequence of Lemma 2.5.1 (a).

2.6.3 Smooth nested Monte Carlo

When the function f is smooth, namely $\mathcal{C}^{1+\rho}(\mathbf{R}, \mathbf{R})$, $\rho \in (0, 1]$ (f' is ρ-Hölder), a variant of the former multilevel nested estimator has been used in [BHR15] (see also [Gil15]) to improve the strong rate of convergence in order to attain the asymptotically unbiased

setting $\beta > 1$ in the condition (SE_β). A root M being given, the idea is to replace in the successive refined levels the difference $Y_{\frac{h}{M}} - Y_h$ (where $h = \frac{1}{K}$, $K \in K_0\mathbf{N}^*$) in the ML2R et MLMC estimators by

$$Y_{h,\frac{h}{M}} := f\left(\frac{1}{MK}\sum_{k=1}^{MK} F(\xi_k, Y)\right) - \frac{1}{M}\sum_{m=1}^{M} f\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi_{(m-1)K+k}, Y)\right).$$

It is clear that

$$\mathbf{E}\left[Y_{h,\frac{h}{M}}\right] = \mathbf{E}\left[Y_{\frac{h}{M}} - Y_h\right].$$

Computations similar to those carried out in Proposition 2.6.1 yield that, if $X = F(\xi, Y) \in L^{p(1+\rho)}(\mathbf{P})$ for some $p \geqslant 2$, then

$$\left\|Y_{h,\frac{h}{M}}\right\|_p^p \leqslant V_M^{(\rho,p)}\left|h - \frac{h}{M}\right|^{\frac{p}{2}(1+\rho)} = V_M^{(\rho,p)}\left|1 - \frac{1}{M}\right|^{\frac{p}{2}(1+\rho)}|h|^{\frac{p}{2}(1+\rho)}. \tag{2.66}$$

▷ *SLLN* : The first consequence is that the SLLN also holds for these modified estimators along the sequences of RMSE $(\varepsilon_k)_{k\geqslant 1}$ satisfying $\sum_{k\geqslant 1}\varepsilon_k^p < +\infty$ owing to Theorem 2.3.1.

▷ *CLT* : When (2.66) is satisfied with $p = 2$, one derives that $\beta = \frac{p}{2}(1+\rho) = 1 + \rho > 1$ whatever ρ is. Hence, the only requested condition in this setting to obtain a CLT (see Theorem 2.3.2) is the L^2–uniform integrability of $\left(h^{-\frac{\beta}{2}}Y_{h,\frac{h}{M}}\right)_{h\in\mathcal{H}}$, since no sharp rate is needed when $\beta > 1$. Moreover, if (2.66) holds for a $p \in (2, +\infty)$, *i.e.* if $X = F(\xi, Y) \in L^{p(1+\rho)}(\mathbf{P})$ with $p > 2$, then

$$h^{-\frac{\beta}{2}}\left\|Y_{h,\frac{h}{M}}\right\|_p \leqslant V_M^{(\rho,p)^{\frac{1}{p}}}\left|1 - \frac{1}{M}\right|^{\frac{1}{2}(1+\rho)}$$

which in turn ensures the L^2–uniform integrability.

As a final remark, note that if the function f *is convex,* $Y_{h,\frac{h}{M}} \leqslant 0$ so that $\mathbf{E}\left[Y_{\frac{h}{M}}\right] \leqslant \mathbf{E}\left[Y_h\right]$ which in turn implies by an easy induction that $\mathbf{E}\left[Y_0\right] \leqslant \mathbf{E}\left[Y_h\right]$ for every $h \in \mathcal{H}$. A noticeable consequence is that the MLMC estimator has a positive bias.

These results can be extended to locally ρ-Hölder continuous functions with polynomial growth at infinity. For more details and a complete proof we refer to Chapter 4.

2.7 Conclusion

We proved a Strong Law of Large Numbers and a Central Limit Theorem for Multilevel estimators with and without weights and we exhibited two applications : the discretization schemes for Brownian diffusions, where we retrieve and slightly extend a result by Ben Alaya and Kebaier in [BAK15], and simulation for nested Monte Carlo, as mentioned in the Multilevel framework by Lemaire and Pagès in [LP17]. The Strong Law of Large Numbers is essentially a consequence of the strong error assumption (SE_β) (or of its reinforced version (2.12)), and the independence od the estimator levels. The understanding of the behaviour of the weights $(\mathbf{W}_r^R)_{r=1,\dots,R}$ in the Multilevel Richardson Romberg estimator is also crucial at this stage, as it is for the proof of the Central Limit Theorem which follows. Under some additional assumptions of L^2-uniform integrability, both the weighted and the standard Multilevel estimators satisfy a Central limit Theorem at rate ε as the quadratic error ε goes to 0. We distinguish between two cases, depending on the value of the strong

error rate β. When $\beta > 1$, both the first coarse level and the successive fine correcting levels contribute to the asymptotic variance of the estimator, whereas when $\beta \in (0, 1]$ the asymptotic variance contains only the contribution of the correcting levels. With the choice of optimal parameters made in Tables 2.1 and 2.2, the Standard Multilevel Monte Carlo estimator has a bias (which is bounded), whereas the weighted Multilevel Monte Carlo estimator is asymptotically without bias, hence we can build exact confidence intervals for the weighted Multilevel Monte Carlo estimator.

Chapitre 3

Antithetic Multilevel for discretization schemes

We give here the methodology for Multilevel Richardson Romberg estimators in the diffusion case, for three types of antithetic schemes, Giles-Szpruch, Ninomiya-Victoir and Giles-Szpruch-Ninomiya-Victoir. This is the counterpart of the work of Al Gerbi, Jourdain and Clément in [AGJC16], which treats the MLMC case. The calibration of the parameters is adjusted for ML2R, starting from the same hypothesis of [AGJC16] and taking into account the antithetic version of the schemes in the computation of the cost.

3.1 General framework

We set $Y_0 = f(X_T)$, where $f : \mathbf{R}^d \to \mathbf{R}$ is a payoff function and X_T is the solution, at time $T \in \mathbf{R}_+^*$, to a multi-dimensional stochastic differential equation of the form

$$\begin{cases} dX_t = b(X_t)dt + \sum_{j=1}^q \sigma^j(X_t)dW_t^j, \quad t \in [0,T], \\ X_0 = x. \end{cases} \tag{3.1}$$

Here, x is the initial condition, $W = (W^1, \dots, W^q)$ is a $q-$dimensional standard Brownian motion, $b : \mathbf{R}^d \to \mathbf{R}^d$ is the drift coefficient and $\sigma^j : \mathbf{R}^d \to \mathbf{R}^d$, $j \in \{1, \dots, q\}$ are the diffusion coefficients.

We want to compute $I_0 = \mathbf{E}[Y_0]$ and we assume that we can approximate the variable Y_0 by choosing a discretization scheme and by discretizing the stochastic differential equation with a time step $h = T/n$, $n \in \mathbf{N}^*$. The most famous discretization schemes is undoubtedly the Euler scheme

$$\begin{cases} X^{EU}_{t^n_{k+1}} = X^{EU}_{t^n_k} + b(X^{EU}_{t^n_k})(t^n_{k+1} - t^n_k) + \sum_{j=1}^q \sigma^j(X^{EU}_{t^n_k})\Delta W^j_{t^n_{k+1}}, \\ X^{EU}_{t_0} = x, \end{cases} \tag{3.2}$$

where $t^n_{k+1} - t^n_k = h$ and $\Delta^j_{t^n_{k+1}} = W_{t^n_{k+1}} - W^j_{t^n_k}$ with $k = 0, \dots, n-1$ and $j = 1, \dots, q$. For a smooth payoff f, it is a well-known result that $Y_h = f(X_T^{EU})$ approximates Y_0 in a weak and in a strong way, satisfying assumption (WE_α) with $\alpha = 1$ and (SE_β) with $\beta = 1$.

We will work with three schemes, the antithetic Giles-Szpruch, the antithetic Ninomiya-Victoir and the combination of these two schemes, the Giles-Szpruch-Ninomiya-Victoir scheme, which we will describe in what follows.

Since an antithetic scheme is constructed by swapping two successive Brownian motions, throughout this chapter we set once for all $M = 2$.

3.2 Antithetic schemes

We give in this Section a general description of antithetic schemes and we recall the results that brings to an assumption (1.28) with $\beta \in (1,2]$ for these schemes. Moreover we describe how to apply the antithetic schemes to Multilevel estimators.

3.2.1 The antithetic scheme for Brownian diffusions : definition and results

We consider a standard Brownian diffusion *SDE* with drift b and diffusion coefficient σ driven by a q-dimensional standard Brownian motion W defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with $q \geqslant 2$ and where the components of $W = (W^1, \dots, W^q)$ are uncorrelated. In such a framework, we saw that Milstein scheme cannot be simulated efficiently due to the presence of Lévy areas induced by the rectangular terms coming out in the second order term of the scheme. This seems to make impossible to reach the unbiased setting "$\beta > 1$" since the Euler scheme corresponds as we saw above to the critical case $\beta = 1$ (scheme of order $\frac{1}{2}$).

First we introduce a truncated Milstein scheme with step $h = \frac{T}{n}$ where the Lévy areas are replaced by their half-sum

$$\begin{cases} \bar{X}^n_{t^n_{k+1}} = \bar{X}^n_{t^n_k} + h\, b\big(t^n_k, \bar{X}^n_{t^n_k}\big) + \sqrt{h}\, \sigma\big(t^n_k, \bar{X}^n_{t^n_k}\big) + \displaystyle\sum_{1 \leqslant i \leqslant j \leqslant q} \bar{\sigma}_{ij}(\bar{X}^n_{t^n_k})\big(\Delta W^i_{t^n_{k+1}} \Delta W^j_{t^n_{k+1}} - \mathbf{1}_{\{i=j\}} h\big), \\ \bar{X}^n_0 = X_0, \end{cases}$$

with $k = 0 : n-1$, $\Delta W_{t^n_{k+1}} = W_{t^n_{k+1}} - W_{t^n_k}$, $t^n_k = \frac{kT}{n}$, $k = 0, \dots, n$ and

$$\bar{\sigma}_{ij}(x) = \frac{1}{2}\big(\partial\sigma_{\cdot i}\, \sigma_{\cdot j} + \partial\sigma_{\cdot j}\, \sigma_{\cdot i}\big),\ 1 \leqslant i < j \leqslant q, \quad \text{and} \quad \bar{\sigma}_{ii}(x) = \frac{1}{2}\partial\sigma_{\cdot i}\, \sigma_{\cdot i}(x),$$

where the tensor $\partial\sigma_{\cdot i}\, \sigma_{\cdot j}$ is defined by

$$\forall\, x = (x_1, \dots, x_d) \in \mathbf{R}^d, \qquad \partial\sigma_{\cdot i}\, \sigma_{\cdot j}(x) := \sum_{\ell=1}^{d} \frac{\partial \sigma_{\cdot i}}{\partial x_\ell}(x)\sigma_{\ell j}(x) \in \mathbf{R}^d. \tag{3.3}$$

This scheme can clearly be simulated, but the analysis of its strong convergence rate, *e.g.* in L^2 when $X_0 \in L^2$, shows a behavior quite similar to the Euler scheme, *i.e.*

$$\Big\| \max_{k=0,\dots,n} |\bar{X}^n_{t^n_k} - X_{t^n_k}| \Big\|_2 \leqslant C_{b,\sigma,T} \sqrt{\frac{T}{n}} \big(1 + \|X_0\|_2\big),$$

if b and σ are $\mathcal{C}^1_{\mathrm{Lip}}$. In terms of weak error it also behaves like the Euler scheme : under similar smoothness assumptions on b, σ and f, or ellipticity on σ, it satisfies a first order expansion in $h = \frac{T}{n}$: $\mathbf{E}\, f(X_T) - \mathbf{E}\, f(\bar{X}^n_T) = c_1 h + O(h^2)$ *i.e.* $(WE)^1_1$ (see [GS14]). Under additional assumptions, it is expected that a higher order weak error expansion $(WE)^1_R$ holds true.

Rather than trying to search for a higher order simulatable scheme, the idea introduced in [GS14] is to combine several such truncated Milstein schemes at different scales in order to make the fine level i behave *as if* the scheme satisfies $(SE)_\beta$ for a $\beta > 1$. To achieve that, the fine scheme of the level is duplicated into two fine schemes with the same step but based on *swapped* Brownian increments. Let us be more specific : the above scheme with

step $h = \frac{T}{n}$ can be described as an homogeneous Markov chain associated to the mapping $\widetilde{\mathcal{M}} = \widetilde{\mathcal{M}}_{b,\sigma} : \mathbf{R}^d \times \mathbf{R}^q \to \mathbf{R}^d$ as follows

$$\bar{X}^n_{t^n_{k+1}} = \widetilde{\mathcal{M}}\big(\bar{X}^n_{t^n_k}, h, \Delta W_{t^n_{k+1}}\big),\ k = 0, \dots, n-1,\ \bar{X}^n_0 = X_0.$$

We consider a first scheme $\big(\bar{X}^{2n,[1]}_{t^{2n}_k}\big)_{k=0,\dots,2n}$ with step $\frac{h}{2} = \frac{T}{2n}$ which reads on two times steps

$$\bar{X}^{2n,[1]}_{t^{2n}_{2k+1}} = \widetilde{\mathcal{M}}\Big(\bar{X}^{2n,[1]}_{t^{2n}_{2k}}, \frac{h}{2}, \Delta W_{t^{2n}_{2k+1}}\Big), \quad \bar{X}^{2n,[1]}_{t^{2n}_{2(k+1)}} = \widetilde{\mathcal{M}}\Big(\bar{X}^{2n,[1]}_{t^{2n}_{2k+1}}, \frac{h}{2}, \Delta W_{t^{2n}_{2(k+1)}}\Big),\ k = 0, \dots, n-1.$$

We consider a second scheme with step $\frac{T}{2n}$, denoted by $\big(\bar{X}^{2n,[2]}_{t^{2n}_k}\big)_{k=0,\dots,2n}$, identical to the first one, except that the Brownian increments are *pairwise swapped i.e.* the increments $, \Delta W_{t^{2n}_{2k+1}}$ and $\Delta W_{t^{2n}_{2(k+1)}}$ are inverted. It reads

$$\bar{X}^{2n,[2]}_{t^{2n}_{2k+1}} = \widetilde{\mathcal{M}}\Big(\bar{X}^{2n,[2]}_{t^{2n}_{2k}}, \frac{h}{2}, \Delta W_{t^{2n}_{2(k+1)}}\Big), \quad \bar{X}^{2n,[2]}_{t^{2n}_{2(k+1)}} = \widetilde{\mathcal{M}}\Big(\bar{X}^{2n,[2]}_{t^{2n}_{2k+1}}, \frac{h}{2}, \Delta W_{t^{2n}_{2k+1}}\Big),\ k = 0, \dots, n-1.$$

The following Theorem established in [GS14] in the autonomous case ($b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$) makes precise the fact that it produces a meta-scheme satisfying $(SE)_\beta$ with $\beta = 2$ (and a root $M = 2$) in a Multilevel scheme.

Theorem 3.2.1. (*a*) *Assume* $b \in \mathcal{C}^2(\mathbf{R}^d, \mathbf{R}^d)$, $\sigma \in \mathcal{C}^2(\mathbf{R}^d, \mathcal{M}(d,q,\mathbf{R}))$ *and that the existing partial derivatives of* b, σ, $\bar{\sigma}$ *(up to order* 2*) are all bounded. Let* $f \in \mathcal{C}^2(\mathbf{R}^d, \mathbf{R})$ *with bounded existing partial derivatives (up to order* 2*) as well. Then, there exists a real constant* $C_{b,\sigma,T} > 0$

$$\Big\| f\big(\bar{X}^n_T\big) - \frac{1}{2}\big(f(\bar{X}^{2n,[1]}_T) + f(\bar{X}^{2n,[2]}_T)\big)\Big\|_2 \leqslant C_{b,\sigma,T}\frac{T}{n}\big(1 + \|X_0\|_2\big), \tag{3.4}$$

i.e. $(SE)_\beta$ *is satisfied with* $\beta = 2$.

(*b*) *Under the above assumptions on* b *and* σ, *assume that* f *is Lipschitz continuous on* $\mathbf{R}^d$, *that its first two order partial derivatives exist outside a Lebesgue negligible set* $\mathcal{N}_0$ *of* $\mathbf{R}^d$ *and are uniformly bounded. Assume that the diffusion* $(X_t)_{t\in[0,T]}$ *satisfies that*

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon^{-1}\, \mathbf{P}\big(\min_{z\in\mathcal{N}_0} |X_T - z| \geqslant \varepsilon\big) < +\infty.$$

then

$$\Big\| f\big(\bar{X}^n_T\big) - \frac{1}{2}\big(f(\bar{X}^{2n,[1]}_T) + f(\bar{X}^{2n,[2]}_T)\big)\Big\|_2 \leqslant C_{b,\sigma,T,\eta}\left(\frac{T}{n}\right)^{\frac{3}{4}-\eta}\big(1 + \|X_0\|_2\big), \tag{3.5}$$

for every $\eta \in (0, \frac{3}{4})$ *i.e.* $(SE)_\beta$ *is satisfied for every* $\beta \in (0, \frac{3}{2})$.

For proofs we refer to Theorems 4.10 and 5.2 in [GS14].

3.2.2 Application to Multilevel estimators with root $M = 2$

The principle is rather simple at this stage. Let us deal with the simplest case of a vanilla payoff (or 1-marginal) $Y_0 = f(X_T)$ and assume that f is such that $(SE)_\beta$ is satisfied for some $\beta \in (1, 2]$.

▷ On the coarse (first) level $i = 1$, we simply set $Y_h^{(1)} = f\big(\bar{X}_T^{n,(1)}\big)$ where the truncated Milstein scheme is driven by a q-dimensional Brownian motion $W^{(1)}$. Keep in mind that this scheme is of order $\frac{1}{2}$ so that all the levels are not ruled by the same β.

▷ At each fine level $i \geqslant 2$, we replace the regular difference $Y^{(i)}_{\frac{h}{n_i}} - Y^{(i)}_{\frac{h}{n_{i-1}}}$ (with $n_i = 2^{i-1}$ and $h \in \mathcal{H}$), by

$$\frac{1}{2}\left(Y^{[1],(i)}_{\frac{h}{n_i}} + Y^{[2],(i)}_{\frac{h}{n_i}}\right) - Y^{(i)}_{\frac{h}{n_{i-1}}},$$

where $Y^{(i)}_{\frac{h}{n_{i-1}}} = f\big(\bar{X}_T^{nM^{i-2},(i)}\big)$, $Y^{[r],(i)}_{\frac{h}{n_i}} = f\big(\bar{X}_T^{nM^{i-1},[r],(i)}\big)$, $r = 1, 2$. At each level the three rough antithetic Milstein schemes are driven by the same q-dimensional Brownian motions $W^{(i)}$, $i = 1, \ldots, R$. Owing to Theorem 3.2.1, it is clear that, under appropriate assumptions,

$$\left\| \frac{1}{2}\left(Y^{[1],(i)}_{\frac{h}{n_i}} + Y^{[2],(i)}_{\frac{h}{n_i}}\right) - Y^{(i)}_{\frac{h}{n_{i-1}}} \right\|_2^2 \leqslant V_2 \left(\frac{h}{n_i}\right)^{\beta}.$$

As a second step, the optimization of the simulation parameter can be carried following the lines of the case $\alpha = 1$ and $\beta = \frac{3}{2}$ or $\beta = 2$. However the first level is ruled at a strong rate $\beta = 1$ which slightly modifies the definition of the sample allocations q_i in Tables 1.3 and 1.2. Moreover, the complexity of fine level K_i is now of the form $\kappa(2n_i + n_{i-1})/h$ instead of $\kappa(n_i + n_{i-1})/h$, which also has to be taken into account in Tables 1.3 and 1.2.

We will work with three antithetic schemes, inspired by the work of Al Gerbi, Jourdian, Clément in [AGJC16]. We give the detail of these three schemes in the following Sections.

3.3 Antithetic Giles-Szpruch scheme

The Giles-Szpruch scheme is a modified version of the Milstein scheme introduced in [GS14]. It is defined as follows

$$\begin{cases} X^{GS}_{t_{k+1}} = X^{GS}_{t_k} + b(X^{GS}_{t_k})(t_{k+1} - t_k) + \displaystyle\sum_{j=1}^{q} \sigma^j(X^{GS}_{t_k}) \Delta W^j_{t_{k+1}} \\ \qquad\qquad + \dfrac{1}{2} \displaystyle\sum_{j,m=1}^{q} \partial\sigma^j \sigma^m(X^{GS}_{t_k}) \left(\Delta W^j_{t_{k+1}} \Delta W^m_{t_{k+1}} - 1_{\{j=m\}} h\right), \\ X^{GS}_{t_0} = x. \end{cases} \tag{3.6}$$

For a smooth payoff, this scheme provides a weak convergence at a rate h^α with $\alpha = 1$ and a strong convergence rate h^β with $\beta = 1$. It is possible to obtain a strong convergence with $\beta = 2$, combining this scheme with its antithetic version $\widetilde{X}^{GS}_T$, which is based on swapping the Brownian increments of X^{GS}_T. More precisely, if the regular Giles-Szpruch scheme on

two successive half time steps reads

$$\begin{cases} X^{GS}_{t_{k+\frac{1}{2}}} = X^{GS}_{t_k} + b(X^{GS}_{t_k})(t_{k+\frac{1}{2}} - t_k) + \sum_{j=1}^{q} \sigma^j(X^{GS}_{t_k})\Delta W^j_{t_{k+\frac{1}{2}}} \\ \qquad + \frac{1}{2}\sum_{j,m=1}^{q} \partial\sigma^j\sigma^m(X^{GS}_{t_k})\left(\Delta W^j_{t_{k+\frac{1}{2}}}\Delta W^m_{t_{k+\frac{1}{2}}} - 1_{\{j=m\}}(t_{k+\frac{1}{2}} - t_k)\right), \\ X^{GS}_{t_{k+1}} = X^{GS}_{t_{k+\frac{1}{2}}} + b(X^{GS}_{t_{k+\frac{1}{2}}})(t_{k+1} - t_{k+\frac{1}{2}}) + \sum_{j=1}^{q} \sigma^j(X^{GS}_{t_{k+\frac{1}{2}}})\Delta W^j_{t_{k+1}} \\ \qquad + \frac{1}{2}\sum_{j,m=1}^{q} \partial\sigma^j\sigma^m(X^{GS}_{t_{k+\frac{1}{2}}})\left(\Delta W^j_{t_{k+1}}\Delta W^m_{t_{k+1}} - 1_{\{j=m\}}(t_{k+1} - t_{k+\frac{1}{2}})\right), \end{cases} \tag{3.7}$$

with $\Delta W^j_{t_{k+\frac{1}{2}}} = W^j_{t_{k+\frac{1}{2}}} - W^j_{t_k}$ and $\Delta W^j_{t_{k+1}} = W^j_{t_{k+1}} - W^j_{t_{k+\frac{1}{2}}}$, then the antithetic scheme is obtained with the same formulas, swapping the Brownian increments as follows

$$\begin{cases} \widetilde{X}^{GS}_{t_{k+\frac{1}{2}}} = \widetilde{X}^{GS}_{t_k} + b(\widetilde{X}^{GS}_{t_k})(t_{k+\frac{1}{2}} - t_k) + \sum_{j=1}^{q} \sigma^j(\widetilde{X}^{GS}_{t_k})\Delta W^j_{t_{k+1}} \\ \qquad + \frac{1}{2}\sum_{j,m=1}^{q} \partial\sigma^j\sigma^m(\widetilde{X}^{GS}_{t_k})\left(\Delta W^j_{t_{k+1}}\Delta W^m_{t_{k+1}} - 1_{\{j=m\}}(t_{k+\frac{1}{2}} - t_k)\right), \\ \widetilde{X}^{GS}_{t_{k+1}} = \widetilde{X}^{GS}_{t_{k+\frac{1}{2}}} + b(\widetilde{X}^{GS}_{t_{k+\frac{1}{2}}})(t_{k+1} - t_{k+\frac{1}{2}}) + \sum_{j=1}^{q} \sigma^j(\widetilde{X}^{GS}_{t_{k+\frac{1}{2}}})\Delta W^j_{t_{k+\frac{1}{2}}} \\ \qquad + \frac{1}{2}\sum_{j,m=1}^{q} \partial\sigma^j\sigma^m(\widetilde{X}^{GS}_{t_{k+\frac{1}{2}}})\left(\Delta W^j_{t_{k+\frac{1}{2}}}\Delta W^m_{t_{k+\frac{1}{2}}} - 1_{\{j=m\}}(t_{k+1} - t_{k+\frac{1}{2}})\right). \end{cases} \tag{3.8}$$

Let $\mathbf{h} = \frac{T}{n}$, $n \in \mathbf{N}^*$. At the fine level $j \geq 2$ coexist a coarse discretization scheme $X_T^{GS,2^{j-2}}$ with time step $h_{j-1} = \frac{T}{n2^{j-2}}$ and two fine schemes associated to the time step $h_j = \frac{T}{n2^{j-1}}$, $X_T^{GS,2^j}$ and its swapped counterpart $\widetilde{X}_T^{GS,2^j}$. Using the notations introduced in 1.2, we define

$$Y^f_{h_j} := \frac{1}{2}\left(f\left(\widetilde{X}_T^{GS,2^{j-1}}\right) + f\left(X_T^{GS,2^{j-1}}\right)\right) \quad \text{and} \quad Y^c_{h_{j-1}} := f\left(X_T^{GS,2^{j-2}}\right). \tag{3.9}$$

We notice that $\mathbf{E}\left[Y^f_{h_j}\right] = \mathbf{E}\left[Y^c_{h_j}\right]$. A weighted Richardson-Romberg Monte Carlo estimator for the antithetic Giles-Szpruch scheme writes

$$I^N_\pi = \sum_{j=1}^{R} \frac{\mathbf{W}^R_j}{N_j} \sum_{k=1}^{N_j} Z_j^{GS,(k)}$$

where the first level is given by

$$Z_1^{GS} = Y^c_{h_1} = f\left(X_T^{GS,1}\right), \tag{3.10}$$

and the following correction levels read

$$Z_j^{GS} = Y^f_{h_j} - Y^c_{h_{j-1}} = \frac{1}{2}\left(f\left(\widetilde{X}_T^{GS,2^{j-1}}\right) + f\left(X_T^{GS,2^{j-1}}\right)\right) - f\left(X_T^{GS,2^{j-2}}\right), \quad j = 2, \ldots, R \tag{3.11}$$

with $(Z_j^{GS})_{j=1,\dots,R}$ independent. The cost of a single diffusion step with the Giles-Szpruch scheme is $\kappa_{GS} := \text{Cost}\left(f\left(X_T^{GS,1}\right)\right)h$. We notice that $\kappa_c = \kappa_{GS}$ and $\kappa_f = 2\kappa_c$, and that the cost function is additive, hence, following (1.14) with $g(x,y) = x + y$, we get

$$K_1 = \text{Cost}(Z_1^{GS}) = \kappa_{GS} n_1 h^{-1} \tag{3.12}$$

and

$$K_j = \text{Cost}(Z_j^{GS}) = 2\kappa_c \frac{n_j}{h} + \kappa_c \frac{n_{j-1}}{h} = \kappa_{GS} h^{-1} M^{j-1} \frac{5}{2}, \quad j = 2, \dots, R. \tag{3.13}$$

When the payoff is smooth, Giles and Szpruch established that the strong convergence assumption (1.28) holds with $\beta = 2$. The weak convergence rate remains the same as for the non antithetic Giles Szpruch scheme, with $\alpha = 1$.

We notice that in [AGJC16] $h = T = 1$ and the cost of a Multilevel Monte Carlo estimator reads

$$\text{Cost}(I_\pi^N) = \kappa \sum_{j=1}^{R} N_j \lambda_j 2^{j-1}.$$

Here $\kappa_{GS} = \kappa$ and, with the notations introduced in Chapter 1, the formula to compute λ_j is given by

$$\lambda_1 = 1, \quad \lambda_j = \frac{\kappa_f + \kappa_c/2}{\kappa_{GS}} = \frac{5}{2}, \quad j = 2, \dots, R.$$

3.4 Antithetic Ninomiya-Victoir scheme

Assuming $\mathcal{C}^1$ regularity for the diffusion coefficients, one can write (3.1) in Stratonovich form

$$\begin{cases} dX_t = \sigma^0(X_t)dt + \sum_{j=1}^q \sigma^j(X_t) \circ dW_t^j, \\ X_0 = x, \end{cases} \tag{3.14}$$

where $\sigma^0 = b - \frac{1}{2}\sum_{j=1}^{q} \partial\sigma^j \sigma^i$ and $\partial\sigma^j$ is the Jacobian matrix of σ^j defined as

$$\partial\sigma^j = \left(\left(\partial\sigma^j\right)_{ik}\right)_{i,k\in[1,d]} = \left(\partial_{x_k}\sigma^{ij}\right)_{i,k\in[1,d]}.$$

The Ninomiya-Victoir scheme then is based on a sequence $\eta = (\eta_1, \dots, \eta_N)$ of independent and identically distributed Rademacher random variables, independent of W, as follows :
if $\eta_{k+1} = 1$:

$$X_{t_{k+1}}^{NV,\eta} = \exp\left(\frac{h}{2}\sigma^0\right) \exp\left(\Delta W_{t_{k+1}}^q \sigma^q\right) \dots \exp\left(\Delta W_{t_{k+1}}^1 \sigma^1\right) \exp\left(\frac{h}{2}\sigma^0\right) X_{t_k}^{NV,\eta}, \tag{3.15}$$

if $\eta_{k+1} = -1$:

$$X_{t_{k+1}}^{NV,\eta} = \exp\left(\frac{h}{2}\sigma^0\right) \exp\left(\Delta W_{t_{k+1}}^1 \sigma^1\right) \dots \exp\left(\Delta W_{t_{k+1}}^q \sigma^q\right) \exp\left(\frac{h}{2}\sigma^0\right) X_{t_k}^{NV,\eta} \tag{3.16}$$

and $X_{t_0}^{NV,\eta} = x$.
For a smooth payoff, the scheme $\frac{1}{2}\left(f\left(X_T^{NV,\eta}\right) + f\left(X_T^{NV,-\eta}\right)\right)$ satisfies (WE_α) with $\alpha = 2$

(see [NV08]) and (1.28) with $\beta = 1$ (see [AGJC16]). Let us take the antithetic version of Ninomiya-Victoir scheme and define

$$Y_{h_j}^f := \frac{1}{4}\left(f\left(\widetilde{X}_T^{NV,2^{j-1},\eta}\right) + f\left(\widetilde{X}_T^{NV,2^{j-1},-\eta}\right) + f\left(X_T^{NV,2^{j-1},\eta}\right) + f\left(X_T^{NV,2^{j-1},-\eta}\right)\right),$$

$$Y_{h_{j-1}}^c := \frac{1}{2}\left(f\left(X_T^{NV,2^{j-2},\eta}\right) + f\left(X_T^{NV,2^{j-2},-\eta}\right)\right)$$

where $\widetilde{X}_T^{NV,2^j,\eta}$ is obtained by swapping the Brownian motions of $X_T^{NV,2^j,\eta}$. A weighted Multilevel Monte Carlo estimator writes

$$I_\pi^N = \sum_{j=1}^{R} \frac{\mathbf{W}_j^R}{N_j} \sum_{k=1}^{N_j} Z_j^{NV,(k)},$$

where the first level reads

$$Z_1^{NV} = Y_{h_1}^c = \frac{1}{2}\left(f\left(X_T^{NV,1,\eta}\right) + f\left(X_T^{NV,1,-\eta}\right)\right)$$

and

$$\begin{aligned} Z_j^{NV} =& Y_{h_j}^f - Y_{h_{j-1}}^c \\ =& \frac{1}{4}\left(f\left(\widetilde{X}_T^{NV,2^{j-1},\eta}\right) + f\left(\widetilde{X}_T^{NV,2^{j-1},-\eta}\right) + f\left(X_T^{NV,2^{j-1},\eta}\right) + f\left(X_T^{NV,2^{j-1},-\eta}\right)\right) \\ & - \frac{1}{2}\left(f\left(X_T^{NV,2^{j-2},\eta}\right) + f\left(X_T^{NV,2^{j-2},-\eta}\right)\right), \quad j = 2,\dots,R \end{aligned} \tag{3.17}$$

and the $(Z_j^{NV})_{j=1,\dots,R}$ are independent. If $h^{-1}\kappa_{NV}$ is the cost of a diffusion $f\left(X_T^{NV,1,\eta}\right)$, hence we get $\kappa_c = 2\kappa_{NV}$ and $\kappa_f = 4\kappa_{NV}$. This yields

$$K_1 = 2\kappa_{NV} h^{-1} n_1, \tag{3.18}$$

$$K_j = 4\kappa_{NV} h^{-1} n_j + 2\kappa_{NV} h^{-1} n_{j-1} = 5\kappa_{NV} h^{-1} M^{j-1}, \quad j = 2,\dots,R, \tag{3.19}$$

which corresponds to $\lambda_1 = 2$ and $\lambda_j = \dfrac{\kappa_f + \kappa_c/2}{\kappa_{NV}} = 5$ in [AGJC16].

It is proved in [AGJC16] that this scheme satisfies (1.28) with a strong rate $\beta = 2$. The weak convergence rate remains the same as for the non antithetic Ninomiya Victoir scheme, with $\alpha = 2$.

3.5 Antithetic Giles-Szpruch-Ninomiya-Victoir scheme

The antithetic Giles-Szpruch-Ninomiya-Victoir scheme introduced in [AGJC16] is a combination of these two schemes, which takes the form (1.16) with $Y_{h_j}^f$ and $Y_{h_{j-1}}^c$ defined in (3.9) for $j = 1,\dots,R-1$, and

$$\bar{Y}_{h_R}^f := \frac{1}{4}\left(f\left(\widetilde{X}_T^{NV,2^{R-1},\eta}\right) + f\left(\widetilde{X}_T^{NV,2^{R-1},-\eta}\right) + f\left(X_T^{NV,2^{R-1},\eta}\right) + f\left(X_T^{NV,2^{R-1},-\eta}\right)\right),$$

$$\bar{Y}_{h_{R-1}}^c := f\left(X_T^{GS,2^{R-2}}\right).$$

Hence the Multilevel Monte Carlo estimator associated to this choice writes

$$I_\pi^N = \frac{1}{N_1}\sum_{k=1}^{N_1} Z_1^{GS,(k)} + \sum_{j=2}^{R-1}\frac{1}{N_j}\sum_{k=1}^{N_j} Z_j^{GS,(k)} + \frac{1}{N_R}\sum_{k=1}^{N_R} \bar{Z}_R^{GSNV,(k)}, \tag{3.20}$$

with

$$Z_1^{GS} = Y_{h_1}^c,$$
$$Z_j^{GS} = Y_{h_j}^f - Y_{h_{j-1}}^c, \quad j = 2, \dots, R-1,$$
$$\bar{Z}_R^{GSNV} = \bar{Y}_{h_R}^f - \bar{Y}_{h_{R-1}}^c.$$

The costs K_j, $j = 1 \dots, R-1$ are given by (3.12) and (3.13) and, for the last level,

$$K_R = 4\kappa_{NV} h^{-1} n_R + \kappa_{GS} h^{-1} n_{R-1} = h^{-1} M^{R-1}\left(4\kappa_{NV} + \frac{\kappa_{GS}}{2}\right).$$

If we make the assumption $\kappa_{NV} = \kappa_{GS}$ we obtain

$$K_R = \frac{9}{2} h^{-1} M^{R-1} \kappa_{GS}, \tag{3.21}$$

which corresponds to $\lambda_1 = 1$, $\lambda_j = 5$ for $j = 2, \dots, R-1$, and $\lambda_R = \dfrac{\kappa_f + \kappa_c/2}{\kappa_{NV}} = 9/2$ in [AGJC16].

This estimator satisfies (1.28) with a strong rate $\beta = 2$ (see [AGJC16] for a complete proof). The weak convergence rate is the same as for the antithetic Ninomiya-Victoir scheme, with $\alpha = 2$. Hence, both the weak and the strong rate are the same as the Ninomiya-Victoir scheme, but with the advantage that since the first $R-1$ levels are build up with a Giles-Szpruch scheme, the estimator (3.20) is less expensive in terms of cost.

3.6 Optimal parameters

In [AGJC16] the optimal parameters are chosen following Giles's method and starting from an assumption on the variance

$$\mathrm{Var}(Z_j) = \frac{c_2}{2^{\beta j}} + o\left(\frac{1}{2^{\beta j}}\right). \tag{3.22}$$

The optimal parameters that we can find in [AGJC16] read

$$L^* = \left\lceil \frac{\log_2\left(\sqrt{2}|c_1|\left(\frac{T}{n}\right)^\alpha \varepsilon^{-1}\right)}{\alpha} \right\rceil + 1 \tag{3.23}$$

and

$$N_\ell = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{\mathrm{Var}(Z_\ell)}{\mathrm{Cost}(Z_\ell)}}\left(\sum_{j=1}^{L}\sqrt{\mathrm{Var}(Z_j)\,\mathrm{Cost}(Z_j)}\right)\right\rceil. \tag{3.24}$$

The first level $\mathrm{Var}(Z_1) = \mathrm{Var}(Y_1)$ is estimated as a structural parameter. The intermediate variances are upper bounded following the assumption (3.22), where the constant c_2 is estimated as a structural parameter. For the Ninomiya-Victoir Giles-Szpruch scheme,

once the level L^* has been computed, we also estimate $\mathrm{Var}(Z_{L^*})$ as a structural parameter. Finally we get, for Giles-Szpruch and for Ninomiya-Victoir antithetic schemes

$$N_1 = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{\mathrm{Var}(Y_1)}{K_1}}\left(\sqrt{\mathrm{Var}(Y_1)K_1} + \sum_{j=2}^{L^*}\sqrt{V_1 2^{-\beta j}K_j}\right)\right\rceil,$$

$$N_\ell = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{V_1 2^{-\beta \ell}}{K_1}}\left(\sqrt{\mathrm{Var}(Y_1)K_1} + \sum_{j=2}^{L^*}\sqrt{V_1 2^{-\beta j}K_j}\right)\right\rceil, \quad \ell = 2,\dots,L^*,$$

with K_j given by (3.12) and (3.13) for Giles-Szpruch scheme, and (3.18) and (3.19) for Ninomiya-Victoir scheme. For Giles-Szpruch Ninomiya-Victoir scheme

$$N_1 = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{\mathrm{Var}(Y_1)}{K_1}}\left(\sqrt{\mathrm{Var}(Y_1)K_1} + \sum_{j=2}^{L^*-1}\sqrt{V_1 2^{-\beta j}K_j} + \sqrt{\mathrm{Var}(Z_{L^*})K_{L^*}}\right)\right\rceil,$$

$$N_\ell = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{V_1 2^{-\beta \ell}}{K_1}}\left(\sqrt{\mathrm{Var}(Y_1)K_1} + \sum_{j=2}^{L^*-1}\sqrt{V_1 2^{-\beta j}K_j} + \sqrt{\mathrm{Var}(Z_{L^*})K_{L^*}}\right)\right\rceil,$$

$$N_{L^*} = \left\lceil \frac{2}{\varepsilon^2}\sqrt{\frac{\mathrm{Var}(Z_{L^*})}{K_{L^*}}}\left(\sqrt{\mathrm{Var}(Y_1)K_1} + \sum_{j=2}^{L^*-1}\sqrt{V_1 2^{-\beta j}K_j} + \sqrt{\mathrm{Var}(Z_{L^*})K_{L^*}}\right)\right\rceil.$$

with $\ell = 2,\dots,L^*-1$. In [LP17] the strong convergence assumption was not based on the variance, but on the L^2–error, as we read in (SE_β), hence the optimal parameters given in Tables (2.1) and (2.2) are not adapted to our framework, but must be substituted with those in the Tables (1.2) and (1.3), by replacing $V_1^* = \mathrm{Var}(Y_1)$, $V_j^* = V_1 2^{-\beta j}$ and $V_R^* = \mathrm{Var}(Z_R)$. Here V_1 corresponds to the c_2 in the [AGJC16] framework. Moreover, in the Tables (2.1) and (2.2), we do not take into account the *antitheticity* of the schemes, since those Tables were computed in the simplest case $Z_\ell = f(X_\ell) - f(X_{\ell-1})$. This problem is solved by replacing the good K_j of formulas (3.12), (3.13), (3.18), (3.19), (3.21) in the new Tables.

Hence we obtain, taking $\widetilde{c}_\infty = 1$ and recalling that c_1 is estimated as a structural parameter, the new Tables (3.1) and (3.2), specifically designed for the combination of antithetic schemes and weighted Multilevel.

Remark 3.6.1. *We notice that in the MLMC case, if we set N_j^{LP} the level sizes computed with Table (3.2) and N_j^{AGJC} those computed in [AGJC16], owing to (1.26), we get $N_j^{LP} = \frac{1}{2}\left(1+\frac{1}{2\alpha}\right)N_j^{AGJC}$. Hence, if $\alpha \geq 1/2$, since $N_j^{LP} \leq N_j^{AGJC}$, we expect the estimators computed with the parameters of Tables (3.1) and (3.2) to be more performing than those computed with the formulas (3.23) and (3.24).*

3.7 Practitioner's corner

The first level variance $\mathrm{Var}(Z_1) = \mathrm{Var}(Y_1)$ and the last level variance $\mathrm{Var}(Z_R)$ (only for Ninomiya-Victoir Giles-Szpruch scheme) are estimated as structural parameters, making a Monte Carlo simulation of size $N = 10^6$ on Y_1 and on Z_R, and taking the empirical variance. We recall that the empirical variance of a Monte Carlo estimator is given by

$R(\varepsilon)$	$\left\lceil \frac{1}{2} + \log_2(\mathbf{h}) + \sqrt{\left(\frac{1}{2} + \log_2(\mathbf{h})\right)^2 + 2\frac{\log_2\left(\sqrt{1+4\alpha}/\varepsilon\right)}{\alpha}} \right\rceil$
$h(\varepsilon)$	$\mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} 2^{-\frac{R-1}{2}} \right\rceil$
$q(\varepsilon)$	$q_j = \mu^* \lvert \mathbf{W}_j^R \rvert \sqrt{\frac{V_j^*}{K_j}}\ j = 1, \dots, R;\ \sum_{1 \leqslant j \leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha R}\right) \frac{1}{\mu^* \varepsilon^2} \sum_{j=1}^{R} \lvert \mathbf{W}_j^R \rvert \sqrt{V_j^* K_j}$

TABLE 3.1 – Optimal parameters for ML2R estimator.

$R(\varepsilon)$	$\left\lceil 1 + \log_2\left(\lvert c_1 \rvert^{\frac{1}{\alpha}} \mathbf{h}\right) + \frac{\log_2(\sqrt{1+2\alpha}/\varepsilon)}{\alpha} \right\rceil$
$h(\varepsilon)$	$\mathbf{h}/\left\lceil \mathbf{h}(1+2\alpha)^{\frac{1}{2\alpha}} \lvert c_1 \rvert^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha}} 2^{-(R-1)} \right\rceil$
$q(\varepsilon)$	$q_j = \mu^* \sqrt{\frac{V_j^*}{K_j}}\ j = 1, \dots, R;\ \sum_{1 \leqslant j \leqslant R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha}\right) \frac{1}{\mu^* \varepsilon^2} \sum_{j=1}^{R} \sqrt{V_j^* K_j}$

TABLE 3.2 – Optimal parameters for MLMC estimator.

$\widehat{\mathrm{Var}}(\frac{1}{N}\sum_{n=1}^N X_n) = \left(\sum_{n=1}^N (X_n)^2 - \frac{1}{N}(\sum_{n=1}^N X_n)^2\right)/(N-1)$. The intermediate variances are upper bounded following the assumption (3.22), where the constant $c_2 = V_1$ is estimated as a structural parameter $\widehat{V}_1$, making a Monte Carlo of size N on $Y_{\frac{h}{M}} - Y_h$, taking the empirical variance $\widehat{\mathrm{Var}}(Y_{\frac{h}{M}} - Y_h)$ and setting

$$\widehat{V}_1 = \widehat{\mathrm{Var}}\left(Y_{\frac{h}{M}} - Y_h\right) h^{-\beta} \left|\frac{1}{M} - 1\right|^{-\beta}.$$

Likewise, the parameter c_1 is estimated by $\widehat{c}_1$, computed using the same Monte Carlo simulation on $Y_{\frac{h}{M}} - Y_h$ and taking

$$\widehat{c}_1 = \frac{1}{N}\sum_{n=1}^N \left(Y_{\frac{h}{M}} - Y_h\right)^{(n)} h^{-\alpha} \left|\frac{1}{M^\alpha} - 1\right|^{-1}.$$

For more details on the importance of a good approximation of $\widehat{c}_1$, we refer to Section 2.4.2.

In the simulations, when we have a closed formula for the payoff, *i.e.* we can compute the true value I_0, we can verify that we respect the theoretical L^2–error that we fixed. The theoretical squared L^2–error writes $\mathbf{E}\left[(I_\pi^N - I_0)^2\right] = \mathbf{E}\left[(I_\pi^N)^2\right] + I_0(I_0 - 2\mathbf{E}\left[I_\pi^N\right])$. We will approximate $\mathbf{E}\left[(I_\pi^N)^2\right]$ and $\mathbf{E}\left[I_\pi^N\right]$ making a standard Monte Carlo estimator of size

$L = 100$ on I_π^N and we will display the empirical L^2–error computed with the formula

$$\sqrt{\frac{1}{L}\sum_{\ell=1}^{L}\left(I_\pi^{N,(\ell)}\right)^2 + I_0\left(I_0 - 2\frac{1}{L}\sum_{\ell=1}^{L} I_\pi^{N,(\ell)}\right)}. \tag{3.25}$$

Moreover, we will display the empirical bias given by

$$\frac{1}{L}\sum_{\ell=1}^{L} I_\pi^{N,(\ell)} - I_0 \tag{3.26}$$

and the empirical variance

$$\frac{1}{L-1}\left(\sum_{\ell=1}^{L}\left(I_\pi^{N,(\ell)}\right)^2 - \frac{1}{L}\left(\sum_{\ell=1}^{L} I_\pi^{N,(\ell)}\right)^2\right). \tag{3.27}$$

3.8 Numerical results

We consider the Clark-Cameron SDE with drift and initial conditions $U_0 = S_0 = 0$, defined as follows

$$\begin{cases} dU_t = S_t dW_t^1, \\ dS_t = \mu dt + dW_t^2, \end{cases} \tag{3.28}$$

where $\mu \in \mathbf{R}$. The Giles-Szpruch scheme for this model is given by

$$\begin{cases} U_{t_{k+1}}^{GS} = U_{t_k}^{GS} + S_{t_k}^{GS}\left(W_{t_{k+1}}^1 - W_{t_k}^1\right) + \frac{1}{2}\left(W_{t_{k+1}}^1 - W_{t_k}^1\right)\left(W_{t_{k+1}}^2 - W_{t_k}^2\right), \\ S_{t_{k+1}}^{GS} = S_{t_k}^{GS} + \mu(t_{k+1} - t_k) + \left(W_{t_{k+1}}^2 - W_{t_k}^2\right), \end{cases} \tag{3.29}$$

for $k = 1, \dots, n-1$. Moreover, in the particular case of this model, we have a closed formula also for the Ninomiya-Victoir scheme, which is given by

$$\begin{cases} U_{t_{k+1}}^{NV,\eta} = U_{t_k}^{NV,\eta} + S_{t_k}^{NV,\eta}\left(W_{t_{k+1}}^1 - W_{t_k}^1\right) + \frac{1}{2}\mu(t_{k+1} - t_k)\left(W_{t_{k+1}}^1 - W_{t_k}^1\right) \\ \qquad\qquad + \mathbf{1}_{\eta_{k+1}=1}\left(W_{t_{k+1}}^1 - W_{t_k}^1\right)\left(W_{t_{k+1}}^2 - W_{t_k}^2\right), \\ S_{t_{k+1}}^{NV,\eta} = S_{t_k}^{NV,\eta} + \mu(t_{k+1} - t_k) + \left(W_{t_{k+1}}^2 - W_{t_k}^2\right), \end{cases} \tag{3.30}$$

for $k = 1, \dots, n-1$. For a smooth payoff function, we recall that previous results in [GS14] and [AGJC16] lead to $\alpha = 1$ and $\beta = 2$ for the antithetic Giles-Szpruch scheme, and $\alpha = \beta = 2$ for both the antithetic Ninomiya-Victoir scheme and the coupling between Giles-Szpruch and Ninomiya-Victoir scheme.

Let us take the smooth payoff $f(u, s) = 10\cos(u)$. Owing to some straightforward computations (see Appendix B), the closed formula for this payoff reads

$$\mathbf{E}\left[\cos(\lambda U_t)\right] = \frac{1}{\sqrt{\cosh \lambda t}} e^{-\frac{\mu^2 t}{2}\left(1 - \frac{\tanh \lambda t}{\lambda t}\right)}.$$

With initial conditions $U_0 = S_0 = 0$, final time $T = 1$, $\mu = 1$ and $\lambda = 1$, the closed formula returns the true value $I_0 = \mathbf{E}\left[10\cos(\lambda U_t)\right] = 7.14556$ (we multiplied by the factor 10 in

	Giles-Szpruch	Ninomiya-Victoir	GSNV	Euler
α	1	2	2	1
β	2	2	2	1
c_1	1.65e+00	5.76e-01	5.74e-01	2.77e+00
V_1	1.40e+01	1.17e+01	1.40e+01	2.03e+01
$\mathrm{Var}(Y_1)$	5.35e+00	1.54e+01	5.42e+00	8.82e+00

TABLE 3.3 – Structural parameters for $f(u,s) = 10\cos(u)$.

order to get a good order of magnitude). The estimated structural parameters are given in Table 3.3.

In Tables from C.1 to C.7 we give all the outputs of the execution of 100 calls to each estimator, where the L^2 error, the bias and the variance were computed with formulas (3.25), (3.26) and (3.27).

3.8.1 Optimal parameters for MLMC estimators

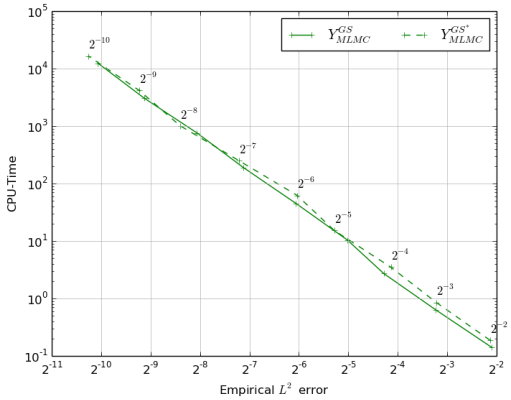
Note that, as expected owing to Remark 3.6.1, the level size N_j computed with the Table (3.2) are smaller than those computed following [AGJC16], hence the CPU-time are smaller, as we can see in Figure (3.1), where $Y_{MLMC}^{(\cdot)^*}$ represents the choice of parameters computed with the Equations (3.23) and (3.24) and where at the same time we can verify that the required accuracy is well fulfilled.

From now on, we will take the optimal parameters for MLMC computed with Table (3.2).

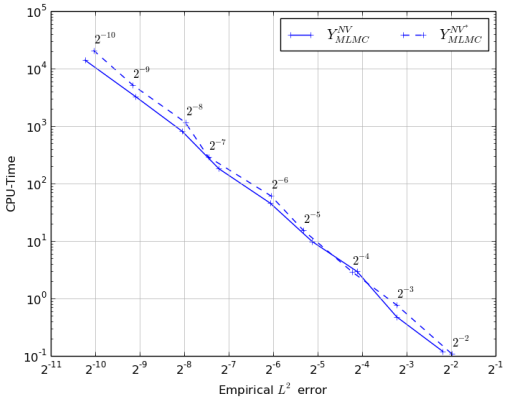
3.8.2 Choice of weak order for ML2R estimator with Ninomiya Victoir scheme

The order of the first term in the bias expansion for the Ninomiya Victoir scheme is 2, hence we set $\alpha = 2$ in the MLMC framework, meaning that $\mathbf{E}[Y_h] = I_0 + c_1 h^2 + o(h^2)$. We recall that an ML2R estimator takes advantage also of the successive terms in the bias expansion, hence we need to know if the term after the first is of order 3 or 4. If it is 4, we can confirm $\alpha = 2$ in the assumption (WE_α), hence $\mathbf{E}[Y_h] = I_0 + \sum_{k=1}^{R} c_k h^{2k} + o(h^{2R})$. But if it is 3, this would mean that assumption (WE_α) holds with $\alpha = 1$ and $c_1 = 0$, hence $\mathbf{E}[Y_h] = I_0 + \sum_{k=2}^{R} c_k h^k + o(h^R)$. Many efforts have been done to improve the order of the first term in the weak error expansion with the Ninomiya Victoir scheme, see for example [OTc12] to find a method to construct weak approximations schemes of order $2m$, $m \in \mathbf{N}$, but to our knowledge there are no results concerning the order of the successive terms. In this particular case, we suspect the weak order α to be 1 with first coefficient $c_1 = 0$. When $c_1 = 0$, as we showed in Section 1.1.2, we need only $R-1$ weights $\widetilde{\mathbf{w}}_r^R$, $r = 1, \dots, R-1$, to kill the terms up to R in the bias polynomial expansion, which we compute with the formula

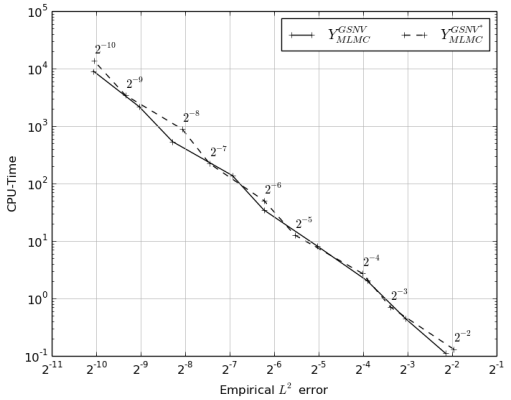
$$\widetilde{\mathbf{w}}_r^R = \frac{n_r^\alpha \, \mathbf{w}_r^{R-1}}{\sum_{j=1}^{R-1} n_j^\alpha \, \mathbf{w}_j^{R-1}} \qquad r = 1, \dots, R-1,$$



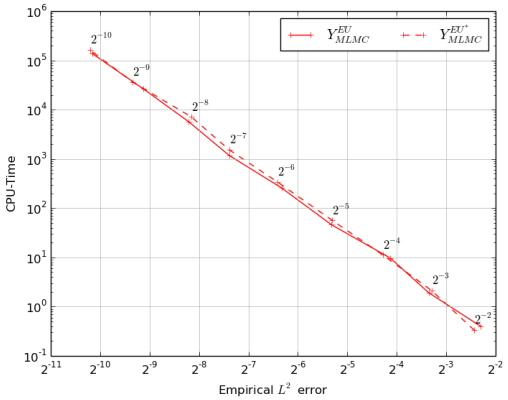
(a) Antithetic Giles Szpruch scheme.



(b) Antithetic Ninomiya Victoir scheme.



(c) Antithetic Giles Szpruch Ninomiya Victoir scheme.



(d) Euler scheme.

FIGURE 3.1 – CPU time vs empirical error for MLMC estimator calibrated with Table 3.2 or with formulas (3.23) and (3.24). Clark-Cameron model with $U_0 = S_0 = 0$, $T = 1$ and $\mu = 1$. Payoff function $f(u, s) = 10\cos(u)$.

where the weights $\mathbf{w}_r^{R-1}$, $r = 1, \dots, R-1$, are the classic weights at order $R-1$ associated to the refiners $n_1 = 1, n_2, \dots, n_{R-1}$.

In what follows, we will take this configuration for the ML2R estimator with Ninomiya Victoir scheme.

3.8.3 ML2R vs MLMC

In Figure 3.2 we can observe the performances of the ML2R estimators, where the accuracy is still fulfilled and the antithetic Giles Szpruch scheme seems to reach better results than antithetic Ninomiya-Victoir scheme and Euler scheme. The gain in using the ML2R estimator instead of MLMC is dramatic with respect to the Euler scheme, as we see in Figure 3.3a, where we show the CPU time ratios with respect to the CPU time of the ML2R estimator with the antithetic Giles Szpruch scheme. There is a significant interest in using the antithetic Giles Szpruch scheme. What should be noticed here is that when using the Euler scheme, the parameter M can vary in order to reach the optimal cost, whereas in the antithetic schemes we forced $M = 2$ (we notice here that the antithetic approach can be used more generally for an arbitrary M, althought it is not yet written up), as we can

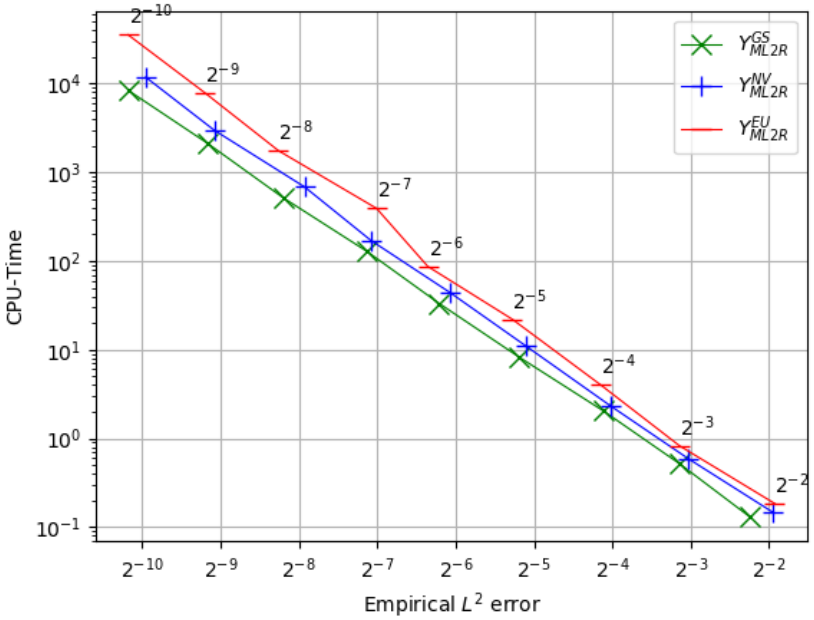
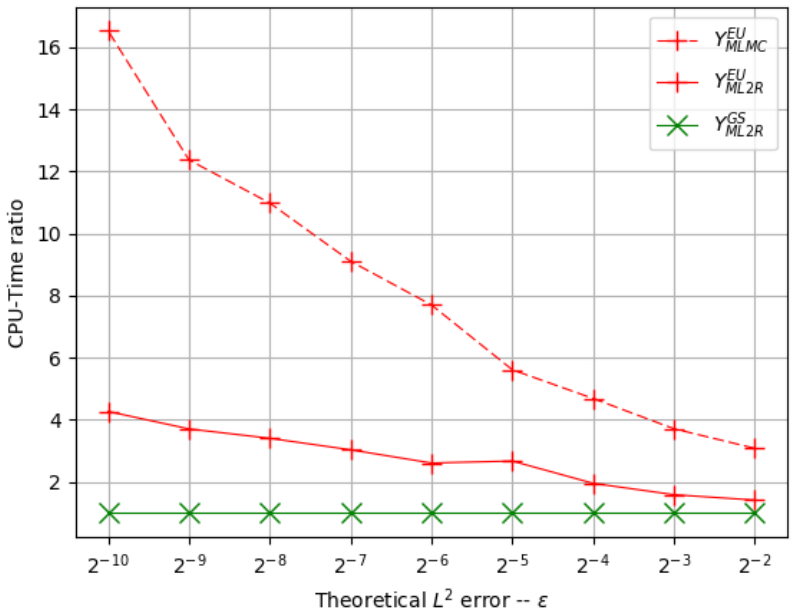


FIGURE 3.2 – ML2R performances with Euler scheme and antithetic schemes.

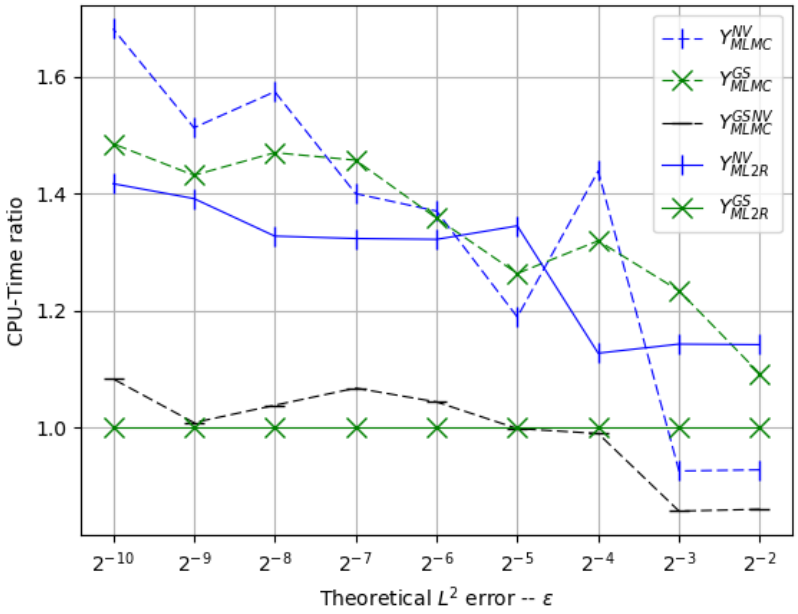
see in Tables C.1 and C.4 in Appendix C.1. Despite the fact that for the antithetic schemes we have the constraint $M = 2$, their better rate $\beta = 2$ still brings a better performance of these estimators with respect to those based on the Euler scheme.

We observe a gain of about 50% when taking the ML2R estimator with antithetic Giles Szpruch scheme with respect to the MLMC schemes with antithetic Giles Szpruch scheme or with antithetic Ninomiya Victoir scheme, whereas there is no significant improvement when taking the ML2R estimator with antithetic Giles Szpruch scheme instead of the MLMC estimator with antithetic Giles Szpruch Ninomiya Victoir scheme, as we see in Figures 3.3b and 3.4.

We notice though that we are in the particular case where we have an explicit expression of the Ninomiya-Victoir scheme, so that we could make the assumption $\kappa_{GS} = \kappa_{NV}$ in (3.21). If we would take a model different from Clark-Cameron, at each time step we would have to solve a set of ODEs, which would significantly increase the value of κ_{NV} and consequently the cost of the estimator. Moreover, when taking an MLMC estimator, we need to estimate the first coefficient c_1 in the bias expansion, whereas for an ML2R estimator such a pre-processing is not necessary. If we would not take this estimated value, as we saw in Section 2.4.1, we would take $c_1 = 1$ and the MLMC estimator would fail owing to an underestimation of the coefficient c_1, as it can observed in Figure 3.5.



(a) Euler scheme (MLMC and ML2R) with respect to antithetic Giles Szpruch scheme (ML2R).



(b) Antithetic schemes (MLMC and ML2R) with respect to antithetic Giles Szpruch scheme (ML2R).

FIGURE 3.3 – CPU time ratios with respect to the ML2R estimator with antithetic Giles Szpruch scheme CPU-time. Clark-Cameron model with $U_0 = S_0 = 0$, $T = 1$ and $\mu = 1$. Payoff function $f(u, s) = 10\cos(u)$.

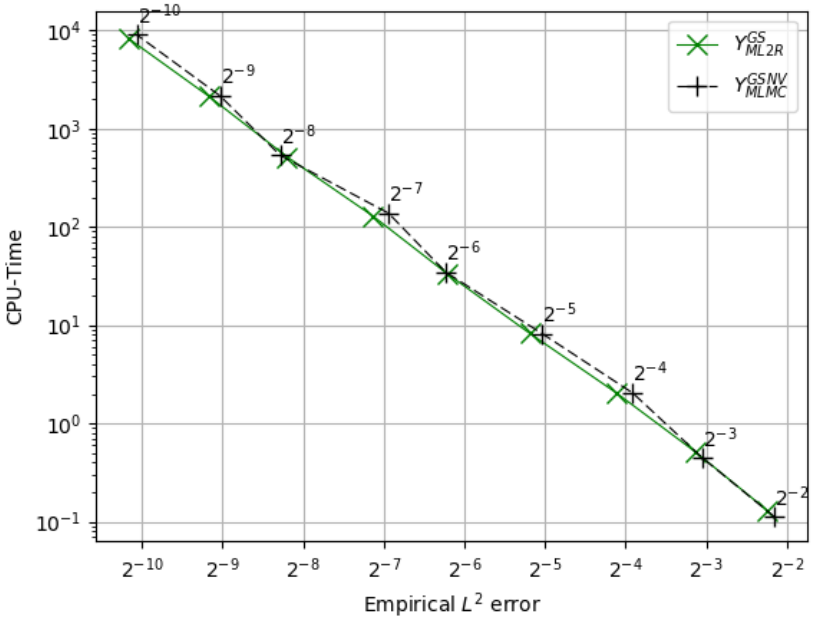


FIGURE 3.4 – ML2R GS vs MLMC GSNV.

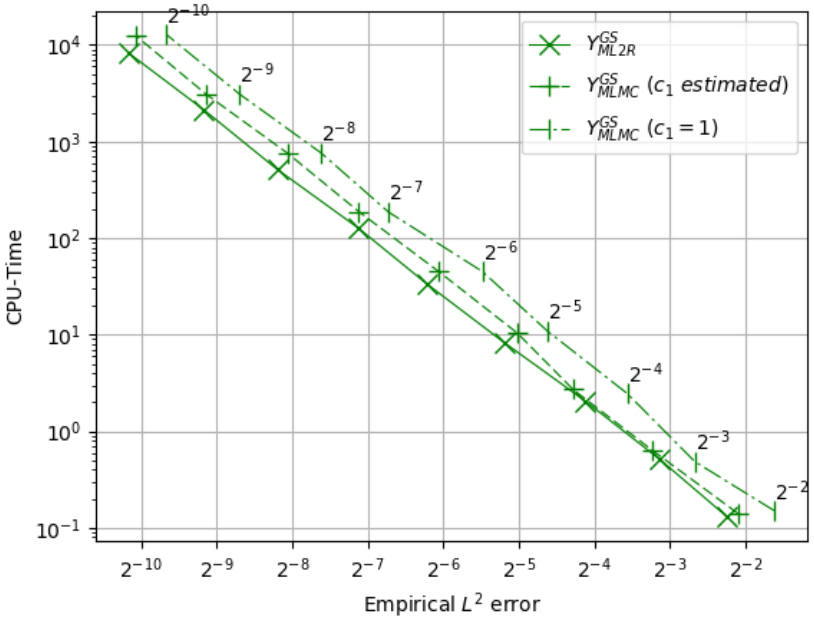


FIGURE 3.5 – Giles-Szpruch with and without c_1 estimation.

Chapitre 4

Nested Monte Carlo

The purpose of the so-called *nested* Monte Carlo method is to compute by simulation quantities of the form

$$\mathbf{E}\left[f\big(\mathbf{E}\left[X \mid Y\right]\big)\right],$$

where (X, Y) is an $\mathbf{R} \times \mathbf{R}^d$-valued couple of random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ satisfying $X \in L^2(\mathbf{P})$ and $f : \mathbf{R} \to \mathbf{R}$ is a specified function such that $f(\mathbf{E}\left[X \mid Y\right]) \in L^2(\mathbf{P})$.

We assume that there exists a Borel function $F : \mathbf{R}^q \times \mathbf{R}^d \to \mathbf{R}$ and random vector $\xi : (\Omega, \mathcal{A}) \to \mathbf{R}^q$ independent of Y such that

$$X = F(\xi, Y).$$

Let us introduce the Borel function $\phi_0 : \mathbf{R}^d \to \mathbf{R}$ defined by $\phi_0(y) = \mathbf{E}\left[F(\xi, y)\right]$ so that one may set $\mathbf{E}\left[F(\xi, Y)|Y\right] = \phi_0(Y)$. Then one has the following representation

$$\mathbf{E}\left[X \mid Y\right] = \phi_0(Y) = \int_{\mathbf{R}^q} F(x, Y)\, \mathbf{P}_\xi(dx).$$

To comply with the multilevel framework, we set $K_0 \in \mathbf{N}^*$ and $\mathcal{H} = \left\{1/K,\, K \in K_0\mathbf{N}^*\right\}$,

$$\bar{X}_0 := \mathbf{E}\left[X \mid Y\right], \quad \bar{X}_h := \frac{1}{K}\sum_{k=1}^{K} F(\xi_k, Y) \text{ with } h = \frac{1}{K} \in \mathcal{H}$$

and $(\xi_k)_{k\geqslant 1}$ is an i.i.d. sequence of random vectors with the same distribution as ξ, defined on $(\Omega, \mathcal{A}, \mathbf{P})$ and independent of Y (up to an enlargement of the probability space if necessary) and

$$Y_0 := f(\bar{X}_0), \quad Y_h := f(\bar{X}_h).$$

We distinguish between two main frameworks, depending on whether or not f is smooth, a classical example of non-smoothness being $f = \mathbf{1}_{(u,s)}$ (see [DL09]). In both cases, our aim is to prove that the nested Monte Carlo estimator satisfies a weak and a strong convergence assumption.

Before getting into the smooth and non smooth case, we give some useful results that will be valid in both frameworks.

4.1 Useful results

Following Comtet [Com74], we introduce the partial Bell polynomials $\mathrm{B}_{n,k}$ for $n \geqslant 1$ and $k = 1, \ldots, n$ defined by

$$\mathrm{B}_{n,k}(x_1, \ldots, x_{n-k+1}) = \sum \frac{n!}{\ell_1! \cdots \ell_{n-k+1}!} \Big(\frac{x_1}{1!}\Big)^{\ell_1} \cdots \Big(\frac{x_{n-k+1}}{(n-k+1)!}\Big)^{\ell_{n-k+1}} \tag{4.1}$$

where the summation takes place over all integers $\ell_1, \dots, \ell_n \geqslant 0$, such that $\ell_1 + 2\ell_2 + \cdots + (n-k+1)\ell_{n-k+1} = n$ and $\ell_1 + \cdots + \ell_{n-k+1} = k$. Note that $d^0 B_{n,k} = k$. The complete Bell polynomials B_n are defined by

$$\mathrm{B}_n(x_1, \dots, x_n) = \sum_{k=1}^{n} \mathrm{B}_{n,k}(x_1, \dots, x_{n-k+1}).$$

The first statement is a formal Taylor expansion with integral remainder of $\mathbf{E}\left[g(\bar{X}_h)\right]$ around $\mathbf{E}\left[g(\bar{X}_0)\right]$, with $g : \mathbf{R} \to \mathbf{R}$ test function.

Lemma 4.1.1 (Taylor expansion)**.** *Let $R \geqslant 0$ and let $g : \mathbf{R} \to \mathbf{C}$ be a $2R+1$ times differentiable function.*

Assume $X \in L^{2R+1}$ and let $\kappa_j(y)$ be the j–th cumulant (a.k.a. semi-invariants) of $F(\xi, y) - \mathbf{E}\left[F(\xi, y)\right]$ for $y \in \mathbf{R}^d$ and $j \in \{1, \dots, R\}$. Let $(\mathrm{B}_{n,k})_{1 \leqslant k \leqslant n}$ be the partial Bell polynomials *defined by (4.1). We then define for $r \geqslant 1$, $r+1 \leqslant n \leqslant 2r$ and $y \in \mathbf{R}^d$,*

$$b_{r,n-r}(y) = \mathrm{B}_{r,n-r}\left(\frac{\kappa_2(y)}{2}, \dots, \frac{\kappa_{2r-n+2}(y)}{2r-n+2}\right).$$

Then

$$\forall h \in \mathcal{H}, \quad \mathbf{E}\left[g(\bar{X}_h)\right] = \mathbf{E}\left[g(\bar{X}_0)\right] + \sum_{r=1}^{2R-1} c(r, (2r+1) \wedge 2R) h^r + \mathcal{R}_{2R+1}, \tag{4.2}$$

with

$$c(r,k) = \frac{1}{r!} \sum_{\ell=r+1}^{k} \mathbf{E}\left[g^{(\ell)}(\bar{X}_0) b_{r,\ell-r}(Y)\right], \ 1 \leqslant r < k \leqslant 2r+1, \tag{4.3}$$

and

$$\mathcal{R}_{2R+1} = \frac{1}{(2R)!} \mathbf{E}\left[\int_0^{\bar{X}_h - \bar{X}_0} g^{(2R+1)}(t + \bar{X}_0)(\bar{X}_h - \bar{X}_0 - t)^{2R} \mathrm{d}t\right]. \tag{4.4}$$

Proof. The case $R = 0$ is trivial, since it is direct application of the fundamental theorem of calculus. Let $h = 1/K$ and let $y \in \mathbf{R}^d$. Let $\tilde{X}_y = F(\xi, y) - \mathbf{E}\left[F(\xi, y)\right]$ and let $E_h(y)$ be the statistical error of the inner Monte Carlo estimator, defined by $E_h(y) = \dfrac{1}{K} \displaystyle\sum_{k=1}^{K} \tilde{X}_y^{(k)}$ where $(\tilde{X}_y^{(k)})_{k \geqslant 1}$ is an *i.i.d.* sequence of copies of $\tilde{X}_y$. For clarity, let us denote $\phi_0(y) = \mathbf{E}\left[F(\xi, y)\right]$.

Let $R \geqslant 1$. The Taylor formula at order $2R$ applied with g gives

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi^{(k)}, y)\right)\right] = g(\phi_0(y)) + \sum_{n=1}^{2R} \frac{g^{(n)}(\phi_0(y))}{n!} \mathbf{E}\left[(E_h(y))^n\right] + \mathcal{R}_{2R+1}(y), \tag{4.5}$$

where $\mathcal{R}_{2R+1}(y) = \frac{1}{(2R)!} \mathbf{E}\left[\int_0^{E_h(y)} g^{(2R+1)}(t + \phi_0(y))(E_h(y) - t)^{2R} \mathrm{d}t\right]$.

The Bell polynomials allow us to explicitly calculate the moments $\mathbf{E}\left[(E_h(y))^n\right]$, $n = 1, \dots, R$ of $E_h(y)$ as follows. Let $\kappa_j(\tilde{X}_y)$, $j = 1, \dots, R$, denote the cumulants (a.k.a. semi-invariants) of $\tilde{X}_y$. Additivity and homogeneity of cumulants give

$$\forall j = 1, \dots, 2R-1, \quad \kappa_j(E_h(y)) = h^{j-1} \kappa_j(\tilde{X}_y).$$

Moments of $E_h(y)$ can be expressed in terms of cumulants using complete Bell polynomials (see [Com74] p.160 Equation(2)) as :

$$\mathbf{E}\left[(E_h(y))^n\right] = \mathrm{B}_n(\kappa_1(E_h(y)), \dots, \kappa_n(E_h(y))).$$

First note that $\kappa_1(\tilde{X}_y) = 0$ so that $\kappa_1(E_h(y)) = 0$. Moreover it follows from the definition (4.1) that $\mathrm{B}_{n,k}$ is k–homogeneous, consequently

$$\mathbf{E}\left[(E_h(y))^n\right] = h^n \sum_{k=1}^{n} h^{-k}\, \mathrm{B}_{n,k}\left(0, \kappa_2(\tilde{X}^y), \dots, \kappa_{n-k+1}(\tilde{X}^y)\right).$$

We derive from (4.1) that $\mathrm{B}_{n,n}(0) = 0$, hence the last term in the above sum is nul and consequently the sum in (4.5) starts from $n = 2$. Note now that

$$\mathrm{B}_{n,k}\left(0, \kappa_2(\tilde{X}^y), \dots, \kappa_{n-k+1}(\tilde{X}^y)\right) = \begin{cases} \frac{n!}{(n-k)!} b_{n-k,k}(y) & \text{if } 1 \leqslant k \leqslant \lceil n/2 \rceil, \\ 0 & \text{if } k > \lceil n/2 \rceil, \end{cases}$$

with $b_{n-k,k}(y) = \mathrm{B}_{n-k,k}\left(\frac{\kappa_2(\tilde{X}^y)}{2}, \dots, \frac{\kappa_{n-2k+2}(\tilde{X}^y)}{n-2k+2}\right)$ which implies that

$$\mathbf{E}\left[(E_h(y))^n\right] = h^n \sum_{k=1}^{\lceil n/2 \rceil} h^{-k} \frac{n!}{(n-k)!} b_{n-k,k}(y). \tag{4.6}$$

Plugging (4.6) in (4.5) gives, since the sum starts at $n = 2$ as mentioned above,

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi^{(k)}, y)\right)\right] = g(\phi_0(y)) + \sum_{n=2}^{2R} g^{(n)}(\phi_0(y)) \sum_{k=1}^{\lceil n/2 \rceil} \frac{h^{n-k}}{(n-k)!} b_{n-k,k}(y) + \mathcal{R}_{2R+1}(y).$$

Letting $r = n-k$ in the above expression, noting that $\lceil n/2 \rceil + \lfloor n/2 \rfloor = n$ and that $\lfloor n/2 \rfloor \leqslant r$ iff $n \leq 2r+1$, on derives by interchanging the sums that

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi^{(k)}, y)\right)\right] = g(\phi_0(y)) + \sum_{r=1}^{2R-1} \frac{h^r}{r!} \left(\sum_{n=r+1}^{(2r+1)\wedge 2R} g^{(n)}(\phi_0(y)) b_{r,n-r}(y)\right) + \mathcal{R}_{2R+1}(y).$$

We conclude by integrating with respect to Y. □

Taking advantage of this expansion we will derive two results. First a bias error expansion for smooth enough payoff functions, in which no regularity is required on the law of $(\bar{X}_0, \bar{X}_h)$ (see Subsection 4.2.1). Conversely a second result will be established relying on the regularity of the distribution of $(\bar{X}_0, \bar{X}_h)$ when the payoff function is not smooth (see Subsection 4.3.1).

As concerns the strong convergence rate of the estimator, elementary computations show that, if $X \in L^2(\mathbf{P})$ (so that $F(\xi, y) \in L^2(\mathbf{P})$ $\mathbf{P}_Y(dy)$-*a.s.*), then

$$\left\|\bar{X}_h - \bar{X}_0\right\|_2^2 = \frac{1}{K}\int \mathbf{P}_Y(dy) \mathrm{var}\big(F(\xi, y)\big) = h\mathbf{E}\left[(F(\xi, Y) - \phi_0(Y))^2\right] \leq h\, \mathrm{var}\big(F(\xi, Y)\big), \tag{4.7}$$

since $\phi_0(Y) = \mathbf{E}\left[F(\xi, Y)|Y\right]$. When $X \in L^p(\mathbf{P})$, we also get a bound in $L^p(\mathbf{P})$, as described in the following Lemma.

Lemma 4.1.2. *Assume $X \in L^p(\mathbf{P})$, $p \geqslant 2$. Then, for every h, $h' \in \mathcal{H}$,*

$$\left\|\bar{X}_h - \bar{X}_{h'}\right\|_p \leqslant 2B_p^{MZ}\left\|X - \mathbf{E}\left[X \,|\, Y\right]\right\|_p |h-h'|^{\frac{1}{2}}. \tag{4.8}$$

Proof. Set $\widetilde{F}(z,y) = F(z,y) - \phi_0(y)$. Assume first that $h, h' \neq 0$, $h' \leqslant h$, and set $K = \frac{1}{h}$, $K' = \frac{1}{h'}$. First note that by Fubini's theorem

$$\left\|\bar{X}_h - \bar{X}_{h'}\right\|_p^p = \int_{\mathbf{R}^d} \mathbf{P}_{_Y}(dy)\mathbf{E}\left[\left|\frac{1}{K}\sum_{k=1}^{K}\widetilde{F}(\xi^k,y) - \frac{1}{K'}\sum_{k=1}^{K'}\widetilde{F}(\xi^k,y)\right|^p\right].$$

Then, for every $y \in \mathbf{R}^d$, it follows from Minkowski's Inequality

$$\left\|\frac{1}{K}\sum_{k=1}^{K}\widetilde{F}(\xi^k,y) - \frac{1}{K'}\sum_{k=1}^{K'}\widetilde{F}(\xi^k,y)\right\|_p \leqslant |h-h'|\left\|\sum_{k=1}^{K}\widetilde{F}(\xi^k,y)\right\|_p + h'\left\|\sum_{k=K+1}^{K'}\widetilde{F}(\xi^k,y)\right\|_p.$$

Applying Marcinkiewicz-Zygmund Inequality to both terms on the right hand side of the above inequality yields

$$\begin{aligned}
&\left\|\frac{1}{K}\sum_{k=1}^{K}\widetilde{F}(\xi^k,y) - \frac{1}{K'}\sum_{k=1}^{K'}\widetilde{F}(\xi^k,y)\right\|_p \\
&\qquad \leqslant |h-h'|B_p^{MZ}\left\|\sum_{k=1}^{K}\widetilde{F}(\xi^k,y)^2\right\|_{\frac{p}{2}}^{\frac{1}{2}} + h'B_p^{MZ}\left\|\sum_{k=K+1}^{K'}\widetilde{F}(\xi^k,y)^2\right\|_{\frac{p}{2}}^{\frac{1}{2}} \\
&\qquad \leqslant |h-h'|B_p^{MZ}K^{\frac{1}{2}}\left\|\widetilde{F}(\xi,y)\right\|_p + h'B_p^{MZ}(K'-K)^{\frac{1}{2}}\left\|\widetilde{F}(\xi,y)\right\|_p,
\end{aligned}$$

where we used again (twice) Minkowski's Inequality in the last line. Finally, for every $y \in \mathbf{R}^d$,

$$\begin{aligned}
\left\|\frac{1}{K}\sum_{k=1}^{K}\widetilde{F}(\xi^k,y) - \frac{1}{K'}\sum_{k=1}^{K'}\widetilde{F}(\xi^k,y)\right\|_p &\leqslant B_p^{MZ}\left\|\widetilde{F}(\xi,y)\right\|_p\left((h-h')\frac{1}{\sqrt{h}} + h'\Big(\frac{1}{h'}-\frac{1}{h}\Big)^{\frac{1}{2}}\right) \\
&= B_p^{MZ}\left\|\widetilde{F}(\xi,y)\right\|_p(h-h')^{\frac{1}{2}}\left(\Big(1-\frac{h'}{h}\Big)^{\frac{1}{2}} + \Big(\frac{h'}{h}\Big)^{\frac{1}{2}}\right) \\
&\leqslant 2\,B_p^{MZ}\left\|\widetilde{F}(\xi,y)\right\|_p(h-h')^{\frac{1}{2}}.
\end{aligned}$$

Plugging this bound in the above equality yields, owing to Minkowski's Inequality and Jensen's Inequality for conditional expectations, the announced result

$$\begin{aligned}
\left\|\bar{X}_h - \bar{X}_{h'}\right\|_p^p &\leqslant (2\,B_p^{MZ})^p\int_{\mathbf{R}^d}\mathbf{P}_{_Y}(dy)\left\|\widetilde{F}(\xi,y)\right\|_p^p(h-h')^{\frac{p}{2}} \\
&= (2\,B_p^{MZ})^p\left\|X - \mathbf{E}\left[X \,|\, Y\right]\right\|_p^p(h-h')^{\frac{p}{2}} \\
&\leqslant (4\,B_p^{MZ})^p\left\|X\right\|_p^p(h-h')^{\frac{p}{2}}.
\end{aligned}$$

□

4.2 Smooth payoff function

We first present the smooth case, where we give a bias error expansion and a strong convergence rate when the payoff function f is smooth.

4.2.1 Weak error

The bias error expansion of the nested Monte Carlo estimator when f is smooth is a consequence of Lemma 4.1.1, as we will see in the proof of the following Proposition.

Proposition 4.2.1 (Bias error (I) : smooth functions)**.** *Let $R \geqslant 1$ and let $f : \mathbf{R} \to \mathbf{R}$ be a $2R+1$ times differentiable payoff function with bounded derivatives $f^{(k)}$, $k = R+1, \dots, 2R+1$. Assume $X \in L^{2R+1}$. Then there exists $c_1, \dots, c_R$ such that*

$$\forall h \in \mathcal{H}, \quad \mathbf{E}\left[f(\bar{X}_h)\right] = \mathbf{E}\left[f(\bar{X}_0)\right] + \sum_{r=1}^{R} c_r h^r + \mathcal{O}(h^{R+1/2}). \tag{4.9}$$

Proof. Applying Lemma 4.1.1 with the function $g = f$ we get

$$\forall h \in \mathcal{H}, \quad \mathbf{E}\left[f(\bar{X}_h)\right] = \mathbf{E}\left[f(\bar{X}_0)\right] + \sum_{r=1}^{R-1} c(r, 2r+1)h^r + c(R, 2R)h^R + \sum_{r=R+1}^{2R-1} c(r, 2R)h^r + \mathcal{R}_{2R+1}, \tag{4.10}$$

with $c(r,k)$ defined in (4.3). Establishing the proposition amounts to proving that the remainder term $\mathcal{R}_{2R+1}$ is well controlled. Using that $f^{(2R+1)}$ is bounded we have

$$\left|\mathcal{R}_{2R+1}\right| \leqslant \frac{\left\|f^{(2R+1)}\right\|_\infty}{(2R+1)!} \mathbf{E}\left[\left|\bar{X}_h - \bar{X}_0\right|^{2R+1}\right].$$

Using successively the Marcinkiewicz-Zygmund Inequality and the Minkowski Inequality for the $L^{R+\frac{1}{2}}(\mathbf{P})$-norm, we get, having in mind that $h = \frac{1}{K}$,

$$\begin{aligned}
\mathbf{E}\left[\left|\frac{1}{K}\sum_{k=1}^{K} F(\xi_k, y) - \phi_0(y)\right|^{2R+1}\right] &\leqslant (B_{2R+1}^{MZ})^{2R+1} h^{2R+1} \mathbf{E}\left[\left|\sum_{k=1}^{K} (F(\xi_k, y) - \phi_0(y))^2\right|^{R+1/2}\right] \\
&\leqslant (B_{2R+1}^{MZ})^{2R+1} h^{R+1/2} \mathbf{E}\left[\left|F(\xi, y) - \phi_0(y)\right|^{2R+1}\right]
\end{aligned}$$

where $B_p^{MZ} = \frac{18 p^{\frac{3}{2}}}{(p-1)^{\frac{1}{2}}}$, $p > 1$ (see [Shi96] p.499). Integrating with respect to $\mathbf{P}_Y$ finally yields

$$\begin{aligned}
\mathbf{E}\left[\left|\bar{X}_h - \bar{X}_0\right|^{2R+1}\right] &\leqslant (B_{2R+1}^{MZ})^{2R+1} h^{R+1/2} \mathbf{E}\left[\left|F(\xi, Y) - \phi_0(Y)\right|^{2R+1}\right] \\
&\leqslant 2^{R+\frac{1}{2}} (B_{2R+1}^{MZ})^{2R+1} h^{R+1/2} \mathbf{E}\left[|X|^{2R+1} + |\mathbf{E}[Y|X]|^{2R+1}\right] \\
&\leqslant 2^{R+\frac{3}{2}} (B_{2R+1}^{MZ})^{2R+1} h^{R+1/2} \mathbf{E}\left[|X|^{2R+1}\right]
\end{aligned} \tag{4.11}$$

so that $\left|\mathcal{R}_{2R+1}\right| = \mathcal{O}(h^{R+1/2})$. □

4.2.2 Strong convergence rate

If we assume that f is Lipschitz continuous, Inequality (4.7) straightforwardly shows that the standard *nested* Monte Carlo satisfies a strong convergence at a rate h^β with $\beta = 1$. When asking for more smoothness, more precisely that f' is ρ–Hölder, we will see that we can build an antithetic version of the *nested* Monte Carlo which attains a strong convergence at a rate h^β with $\beta > 1$. As we saw this corresponds to the optimal unbiased setting in terms of minimization of the computational cost.

4.2.2.1 f Lipschitz continuous

It is a known result that if f is Lipschitz continuous, then the *nested* Monte Carlo satisfies a strong convergence assumption with $\beta = 1$. More precisely the following Proposition holds.

Proposition 4.2.2. *Assume f is Lipschitz continuous. For every $h, h' \in \mathcal{H} \cup \{0\}$,*

$$\|Y_{h'} - Y_h\|_2^2 \leq [f]_{\mathrm{Lip}}^2 (\|X\|_2^2 - \|\mathbf{E}[X|Y]\|_2^2)|h' - h|$$

so that $(Y_h)_{h\in\mathcal{H}}$ satisfies a strong convergence assumption with $\beta = 1$.

This is a straightforward consequence of Inequality (4.7).

4.2.2.2 f' locally Hölder : antithetic approach

When the function f is $\mathcal{C}^{1+\rho}(\mathbf{R}, \mathbf{R})$, $\rho \in (0, 1]$ (f' is ρ-Hölder), a variant of the former Multilevel *nested* estimator has been proposed in [BHR15], [HA12] and [CL12] (see also [Gil15]) to improve the rate of strong convergence in order to attain the asymptotically unbiased setting, with $\beta > 1$. We recall that, except for the antithetic discretization schemes for Brownian diffusions, the root has no reason to be constrained to be $M = 2$. In this Section we will consider general refiners of the form $n_j = M^{j-1}$, $j = 1, \ldots, R$, $M \geq 2$. A root M being given, the idea is to replace in the successive refined levels the difference $Y_{\frac{h}{M}} - Y_h$ (where $h = \frac{1}{K}$, $K \in K_0\mathbf{N}^*$) in the ML2R et MLMC estimators by an antithetic type as follows

$$Y_{h,\frac{h}{M}} := f\left(\frac{1}{MK}\sum_{k=1}^{MK} F(\xi_k, Y)\right) - \frac{1}{M}\sum_{m=1}^{M} f\left(\frac{1}{K}\sum_{k=1}^{K} F(\xi_{(m-1)K+k}, Y)\right).$$

It is clear that $\mathbf{E}\left[Y_{h,\frac{h}{M}}\right] = \mathbf{E}\left[Y_{\frac{h}{M}} - Y_h\right]$.

Proposition 4.2.3. *Assume $X \in L^{p(1+\rho)}$ and f' ρ–Hölder, i.e.*

$$\left|f'(x) - f'(y)\right| \leq \left[f'\right]_\rho |x - y|^\rho, \quad 0 < \rho \leq 1. \tag{4.12}$$

Then $Y_{h,\frac{h}{M}}$ satisfies a strong convergence assumption with $\beta = 1 + \rho > 1$.

Proof. We notice that, owing to Burkholder's Inequality, for all $p \geq 2$, for every sequence $(X_k)_{k\geq 0}$ of random variables centered and *i.i.d.* and $K \in \mathbf{N}$, $K \geq 2$, there exists a positive real constant C_p such that

$$\mathbf{E}\left[\left|\sum_{k=1}^{K} X_k\right|^p\right] \leq C_p \mathbf{E}\left[\left|\sum_{k=1}^{K} X_k^2\right|^{\frac{p}{2}}\right] \leq C_p \left(\sum_{k=1}^{K} \left\|X_k^2\right\|_{\frac{p}{2}}\right)^{\frac{p}{2}} = C_p K^{\frac{p}{2}} \mathbf{E}\left[|X_1|^p\right]. \tag{4.13}$$

We set

$$\bar{X}_{K,m} = \frac{1}{K}\sum_{k=1}^{K} F(\xi_{K(m-1)+k}, Y) \quad \text{and} \quad \bar{X}_{MK} = \frac{1}{M}\sum_{m=1}^{M} \bar{X}_{K,m} = \frac{1}{MK}\sum_{k=1}^{MK} F(\xi_k, Y).$$

The nested multilevel Monte Carlo estimator then reads

$$I_\pi^N(\varepsilon) = \frac{1}{N_1}\sum_{i=1}^{N_1} f(\bar{X}_{K,1}^{(i)}) + \sum_{j=2}^{R}\frac{1}{N_j}\sum_{i=1}^{N_j}\left(f(\bar{X}_{MK}^{(i)}) - \frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}^{(i)})\right),$$

with $(\bar{X}_{K,m}^{(i)})_{i\geq 1}$ independent copies of $\bar{X}_{K,m}$. Owing to Taylor formula, for all $m = 1,\dots,M$, there exists $x_m \in \big(\min(\bar{X}_{K,m}, \bar{X}_{MK}), \max(\bar{X}_{K,m}, \bar{X}_{MK})\big)$ such that

$$f(\bar{X}_{K,m}) = f(\bar{X}_{MK}) + f'(\bar{X}_{MK})(\bar{X}_{K,m} - \bar{X}_{MK}) + \big(f'(x_m) - f'(\bar{X}_{MK})\big)(\bar{X}_{K,m} - \bar{X}_{MK}).$$

Hence, using that $\bar{X}_{K,1}$ is independent of $\big(\bar{X}_{K,m}\big)_{m=2,\dots,M}$,

$$\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) = f(\bar{X}_{MK}) + 0 + \frac{1}{M}\sum_{m=1}^{M}\big(f'(x_m) - f'(\bar{X}_{MK})\big)(\bar{X}_{K,m} - \bar{X}_{MK}). \quad (4.14)$$

We aim at computing $\left\|Y_{h,\frac{h}{M}}\right\|_p = \left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p$, with $p \geq 2$. Owing to the decomposition (4.14), to Minkowski's Inequality and to the ρ–Hölder assumption (4.12), we get

$$\begin{aligned}\left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p &= \left\|\frac{1}{M}\sum_{m=1}^{M}\big(f'(x_m) - f'(\bar{X}_{MK})\big)(\bar{X}_{K,m} - \bar{X}_{MK})\right\|_p \\ &\leq [f']_\rho \frac{1}{M}\sum_{m=1}^{M}\left\||\bar{X}_{K,m} - \bar{X}_{MK}|^{1+\rho}\right\|_p. \quad (4.15)\end{aligned}$$

We first notice by an exchangeability argument that the variables $(\bar{X}_{K,m} - \bar{X}_{MK})_{m=1\dots,M}$ are identically distributed with $\bar{X}_{K,m} - \bar{X}_{MK} \sim \bar{X}_{K,1} - \bar{X}_{MK}$. Moreover we write

$$\bar{X}_{K,1} - \bar{X}_{MK} = \bar{X}_{K,1} - \frac{1}{M}\sum_{m=1}^{M}\bar{X}_{K,m} = \frac{1}{M}\sum_{m=1}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m}) = \frac{1}{M}\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m}).$$

Hence, we get

$$\left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p \leq [f']_\rho \frac{1}{M^{1+\rho}}\left\|\left|\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{1+\rho}\right\|_p. \quad (4.16)$$

Owing to the independence of Y and $(\xi_k)_{k\geq 1}$, we write

$$\begin{aligned}&\mathbf{E}\left[\left|\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{(1+\rho)p}\right] \\ &= \int \mathbf{P}_Y(dy)\mathbf{E}\left[\left|\sum_{m=2}^{M}\left(\frac{1}{K}\sum_{k=1}^{K}F(\xi_k, y) - \frac{1}{K}\sum_{k=1}^{K}F(\xi_{K(m-1)+k}, y)\right)\right|^{(1+\rho)p}\right] \\ &= \int \mathbf{P}_Y(dy)\frac{1}{K^{(1+\rho)p}}\mathbf{E}\left[\left|\sum_{k=1}^{K}\left((M-1)F(\xi_k, y) - \sum_{m=2}^{M}F(\xi_{K(m-1)+k}, y)\right)\right|^{(1+\rho)p}\right].\end{aligned}$$

We notice that, for each fixed $y \in \mathbf{R}^d$, the random variables $(X_k)_{k\geq 1} = \left((M-1)F(\xi_k,y) - \sum_{m=2}^M F(\xi_{K(m-1)+k},y)\right)_{k\geq 1}$ are centered and *i.i.d.*. Moreover $(1+\rho)p \geq 2$ hence, owing to Inequality (4.13), there exists a constant $C_{p,\rho}$ such that

$$\mathbf{E}\left[\left|\sum_{m=2}^M (\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{(1+\rho)p}\right] \leq C_{p,\rho}\frac{1}{K^{\frac{(1+\rho)p}{2}}}\int \mathbf{P}_Y(dy)\mathbf{E}\left[\left|(M-1)F(\xi_1,y) - \sum_{m=2}^M F(\xi_{K(m-1)+1},y)\right|^{(1+\rho)p}\right].$$

Applying twice Minkowski's Inequality yields

$$\begin{aligned}
&\mathbf{E}\left[\left|\sum_{m=2}^M (\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{(1+\rho)p}\right] \\
&\leq C_{p,\rho}\frac{1}{K^{\frac{(1+\rho)p}{2}}}(M-1)^{(1+\rho)p}\int \mathbf{P}_Y(dy)\mathbf{E}\left[|F(\xi_1,y) - F(\xi_{K+1},y)|^{(1+\rho)p}\right] \\
&\leq C_{p,\rho}\frac{1}{K^{\frac{(1+\rho)p}{2}}}(M-1)^{(1+\rho)p}2^{(1+\rho)p}\mathbf{E}\left[|F(\xi_1,Y)|^{(1+\rho)p}\right].
\end{aligned}$$

Plugging this in (4.16) we get

$$\left\|\frac{1}{M}\sum_{m=1}^M f(\bar{X}_{K,m}) - f(\bar{x})\right\|_p^p \leq V_1^{(p,\rho)}\frac{1}{K^{\frac{p(1+\rho)}{2}}},$$

with $V_1^{(p,\rho)} = [f']_\rho^p C_{p,\rho}2^{p(1+\rho)}\left(1-\frac{1}{M}\right)^{(1+\rho)p}\mathbf{E}\left[|F(\xi_1,Y)|^{(1+\rho)p}\right]$. Hence, under the assumption $X \in L^{p(1+\rho)}$, $Y_{h,\frac{h}{M}}$ satisfies the L^p version of the strong convergence assumption with $\beta = 1+\rho > 1$. □

If we replace the ρ–Hölder assumption on f' by a locally ρ–Hölder assumption, *i.e.*

$$\left|f'(x) - f'(y)\right| \leq C|x-y|^\rho\left(1+|x|^q+|y|^q\right),$$

then, since $|x_m|^q \leq \max(|\bar{X}_{K,m}|^q, |\bar{X}_{MK}|^q) \leq |\bar{X}_{K,m}|^q + |\bar{X}_{MK}|^q$, Inequality (4.15) must be replaced by

$$\begin{aligned}
\left\|\frac{1}{M}\sum_{m=1}^M f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p &= \left\|\frac{1}{M}\sum_{m=1}^M \left(f'(x_m) - f'(\bar{X}_{MK})\right)(\bar{X}_{K,m} - \bar{X}_{MK})\right\|_p \\
&\leq [f']_\rho\frac{1}{M}\sum_{m=1}^M \left\||\bar{X}_{K,m} - \bar{X}_{MK}|^{1+\rho}(1+|\bar{X}_{K,m}|^q + 2|\bar{X}_{MK}|^q)\right\|_p. \quad (4.17)
\end{aligned}$$

Owing to Hölder's Inequality with $r,s>1$ such that $\frac{1}{r}+\frac{1}{s}=1$ and Minkowski's Inequality, we get

$$\begin{aligned}
&\left\||\bar{X}_{K,m} - \bar{X}_{MK}|^{1+\rho}(1+|\bar{X}_{K,m}|^q + 2|\bar{X}_{MK}|^q)\right\|_p \\
&\qquad \leq \left\||\bar{X}_{K,m} - \bar{X}_{MK}|^{1+\rho}\right\|_{pr}\left(1 + \left\||\bar{X}_{K,m}|^q\right\|_{ps} + 2\left\||\bar{X}_{MK}|^q\right\|_{ps}\right).
\end{aligned}$$

We recall that the variables $(\bar{X}_{K,m})_{m=1,\dots,M}$ are identically distributed, hence Inequality (4.17) yields

$$\left\| \frac{1}{M} \sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK}) \right\|_p \leq [f']_\rho \left\| |\bar{X}_{K,1} - \bar{X}_{MK}|^{1+\rho} \right\|_{pr} \left(1 + \left\| |\bar{X}_{K,1}|^q \right\|_{ps} + 2 \left\| \bar{X}_{MK}|^q \right\|_{ps} \right).$$

The analysis of the term $\left\| |\bar{X}_{K,1} - \bar{X}_{MK}|^{1+\rho} \right\|_{pr}$ does not change, except for the condition $X \in L^{(1+\rho)pr}$. Under the assumption $X \in L^{qps}$, the term $1 + \left\| |\bar{X}_{K,1}|^q \right\|_{ps} + 2 \left\| |\bar{X}_{MK}|^q \right\|_{ps}$ is bounded, since

$$\left\| |\bar{X}_{K,1}|^q \right\|_{ps} = \left\| \left| \frac{1}{K} \sum_{k=1}^{K} X_k \right|^q \right\|_{ps} \leq \left(\|X_1\|_{qps \vee 1} \right)^q.$$

Having in mind that $r = s/(s-1)$, the optimal choice for s which minimizes both $(1+\rho)pr$ and qps is given by $s = (1+\rho+q)/q$ (hence $r = (1+\rho+q)/(1+\rho)$). This leads to the additional condition to $X \in L^{p(1+\rho+q)}$. In conclusion, we obtain that

$$\left\| \frac{1}{M} \sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK}) \right\|_p^p \leq V_1^{(p,q,\rho)} \frac{1}{K^{\frac{p(1+\rho)}{2}}},$$

with $V_1^{(p,q,\rho)} = [f']_\rho^p \left(C_{pr,\rho}^1 C_{pr,\rho}^2 2^{pr(1+\rho)} \right)^{\frac{1}{r}} M^{-\frac{p(1+\rho)}{2}} \left\| |X_1|^{1+\rho} \right\|_{pr}^p \left(1 + 3 \left(\|X_1\|_{qps \vee 1} \right)^q \right)^p$. Hence, under the assumption $X \in L^{p(1+\rho+q)}$, $Y_{h,\frac{h}{M}}$ satisfies the L^p version of the strong convergence assumption with $\beta = 1 + \rho > 1$.

4.3 Smooth density

There are many situations where we need to consider non smooth payoff functions of the type $f = \mathbf{1}_{\{g(\mathbf{E}[X|Y]) \in I\}}$, with $g : \mathbf{R} \to \mathbf{R}$ and $I \subset \mathbf{R}$ interval. Among them we can cite the computation of loss thresholds, *i.e.* when we search, a threshold $q \in \mathbf{R}$ being fixed, for the corresponding $\alpha_q \in [0,1]$ such that

$$1 - \alpha_q = \mathbf{P}(g(\mathbf{E}[X|Y]) \geq q) = \mathbf{E}\left[\mathbf{1}_{\{g(\mathbf{E}[X|Y]) \geq q\}} \right],$$

or the inverse problem, which consists in computing the quantile q_α such that for a fixed $\alpha \in [0,1]$,

$$1 - \alpha = \mathbf{P}(g(\mathbf{E}[X|Y]) \geq q_\alpha) = \mathbf{E}\left[\mathbf{1}_{\{g(\mathbf{E}[X|Y]) \geq q_\alpha\}} \right].$$

Another situation of interest is the approximation of density functions (see the seminal paper of Bally and Talay [BT96a] and [BT96b], treating the law of the Euler scheme for distributions).

The payoff function f being non smooth, the regularity assumptions on f that we needed to prove the weak and the strong convergence of the estimator in the smooth case, will be replaced by some assumptions on the regularity of the density functions, as we detail in the next two Subsections.

4.3.1 Weak error

We recall the notation

$$\bar{X}_0 := \mathbf{E}\left[X \mid Y\right], \quad \bar{X}_h := \frac{1}{K}\sum_{k=1}^{K} F(\xi_k, Y) \text{ with } h = \frac{1}{K} \in \mathcal{H}$$

The following result on the weak error derives from Lemma 4.1.1 and gives a bias error expansion relying on the density of the joint distribution of $(\bar{X}_0, \bar{X}_h)$. More precisely, assume that $(\bar{X}_0, \bar{X}_h - \bar{X}_0)$ is a random vector with smooth density with respect to the Lebesgue measure on $\mathbf{R}^2$. Let $f_{\bar{X}_0}$ be the density of $\bar{X}_0$ and let $f_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}$ be (a regular version of) the conditional density of $\bar{X}_0$ given $\bar{X}_h - \bar{X}_0 = \bar{x}$, so that for all test functions $g : \mathbf{R} \to \mathbf{R}$, $\mathbf{E}\left[g(\bar{X}_0)|\bar{X}_h - \bar{X}_0\right] = \int g(x) f_{\bar{X}_0|\bar{X}_h-\bar{X}_0}(x)dx$. Moreover let $F_{\bar{X}_h}(x)$ and $F_{\bar{X}_0}(x)$ be the distribution functions of $\bar{X}_h$ and $\bar{X}_0$.

Proposition 4.3.1 (Bias error (II) : smooth density)**.** *(a) Let $R \geqslant 0$. Assume that the functions ϕ_0, $f_{\bar{X}_0}$, $f_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}$, $\bar{x} \in \mathbf{R}$, are $2R+1$ times differentiable and $f^{(2R+1)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}$ is continuous. Assume that $\phi_0 : \mathbf{R} \to \mathbf{R}$ is continuous and strictly monotonic, that its derivative ϕ_0' is never 0, hence ϕ_0 is one-to-one, and assume that $X \in L^{2R+1}$.*

Let $\psi_k = f^{(k)}_{\bar{X}_0}/f_{\bar{X}_0}$ and

$$\begin{aligned} P_r(x) &= \frac{1}{f_{\bar{X}_0}(x)} \sum_{\ell=r+1}^{(2r+1)\wedge 2R} (-1)^\ell \left(\left(b_{r,\ell-r} \circ \phi_0^{-1}\right).f_{\bar{X}_0} \right)^{(\ell)}(x) \\ &= \sum_{\ell=r+1}^{(2r+1)\wedge 2R} (-1)^\ell \sum_{k=0}^{\ell} \binom{\ell}{k} \left(b_{r,\ell-r} \circ \phi_0^{-1}\right)^{(\ell-k)}(x)\psi_k(x). \end{aligned} \tag{4.18}$$

(a) If $\sup_{h\in\mathcal{H},\bar{x},x\in\mathbf{R}} |f^{(2R+1)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x)| < +\infty$, then

$$f_{\bar{X}_h}(x) = f_{\bar{X}_0}(x) + f_{\bar{X}_0}(x) \sum_{r=1}^{R} \frac{h^r}{r!} P_r(x) + \mathcal{O}(h^{R+\frac{1}{2}}) \tag{4.19}$$

uniformly with respect to $x \in \mathbf{R}$.

(b) If furthermore $\sup_{h\in\mathcal{H},\bar{x},x\in\mathbf{R}} |f^{(2R)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x)| < +\infty$ and $\lim_{x\to-\infty} f^{(2R)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x) = 0$ for every $\bar{x} \in \mathbf{R}$, then

$$F_{\bar{X}_h}(x) = F_{\bar{X}_0}(x) + \sum_{r=1}^{R} \frac{h^r}{r!} \mathbf{E}\left[P_r(\bar{X}_0)\mathbf{1}_{\{\bar{X}_0 \leqslant x\}}\right] + \mathcal{O}(h^{R+\frac{1}{2}}) \tag{4.20}$$

uniformly with respect to $x \in \mathbf{R}$.

Proof. The case $R = 0$ is trivial, using the expansion (4.2) and the convention $\sum_{r=1}^{0} = 0$. Let $g : \mathbf{R} \to \mathbf{R}$ be an infinitely differentiable function with compact support. We apply the expansion (4.2) to the smooth function g where coefficients $c(r, 2r+1)$, $r = 1, \ldots, R-1$ and $c(r, 2R)$, $r = R, \ldots, 2R-1$ are given by (4.3) and the remainder term $\mathcal{R}^g_{2R+1}$ is given by (4.4).

We first note that, for every $\ell \in \{1, \ldots, 2R\}$,

$$\mathbf{E}\left[g^{(\ell)}(\bar{X}_0)b_{r,\ell-r}(Y)\right] = \mathbf{E}\left[g^{(\ell)}(\bar{X}_0)(b_{r,\ell-r}\circ\phi_0^{-1})(\bar{X}_0)\right] = \int_{\mathbf{R}} g^{(\ell)}(x)\left((b_{r,\ell-r}\circ\phi_0^{-1})f_{\bar{X}_0}\right)(x)\mathrm{d}x.$$

Then, performing successively ℓ integrations by parts yields

$$\mathbf{E}\left[g^{(\ell)}(\bar{X}_0)b_{r,\ell-r}(Y)\right] = (-1)^\ell \int_{\mathbf{R}} g(x)\big((b_{r,\ell-r}\circ\phi_0^{-1})f_{\bar{X}_0}\big)^{(\ell)}(x)\mathrm{d}x.$$

As for the remainder term,

$$\begin{aligned}
\mathcal{R}^g_{2R+1} &= \frac{1}{(2R)!}\mathbf{E}\left[\int_0^{\bar{X}_h-\bar{X}_0} g^{(2R+1)}(t+\bar{X}_0)(\bar{X}_h-\bar{X}_0-t)^{2R}\mathrm{d}t\right]\\
&= \frac{1}{(2R)!}\int_{\mathbf{R}} \mathrm{d}\bar{x}\, f_{\bar{X}_h-\bar{X}_0}(\bar{x})\int_0^{\bar{x}}\left[\int_{\mathbf{R}} g^{(2R+1)}(t+x)f_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}(x)\mathrm{d}x\right](\bar{x}-t)^{2R}\mathrm{d}t\\
&= \frac{1}{(2R)!}\int_{\mathbf{R}} \mathrm{d}\bar{x}\, f_{\bar{X}_h-\bar{X}_0}(\bar{x})\int_0^{\bar{x}} g^{(2R+1)}*\check{f}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}(t)(\bar{x}-t)^{2R}\mathrm{d}t,
\end{aligned}$$

where $\check{f}: u\mapsto f(-u)$ and $*$ denotes standard convolution of functions on $\mathbf{R}$. Now, owing to the regularity assumption made on $f_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}$, we know that $g^{(2R+1)}*\check{f}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}} = g*(\check{f}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}})^{(2R+1)} = -\check{f}^{(2R+1)}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}*g$. It follows from Fubini's Theorem that

$$\begin{aligned}
\mathcal{R}^g_{2R+1} &= \frac{1}{(2R)!}\int_{\mathbf{R}} \mathrm{d}\bar{x} f_{\bar{X}_h-\bar{X}_0}(\bar{x})\int_0^{\bar{x}}\mathrm{d}t\int_{\mathbf{R}}\mathrm{d}x g(x) f^{(2R+1)}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}(x-t)(\bar{x}-t)^{2R}\\
&= -\frac{1}{(2R)!}\int_{\mathbf{R}} \mathrm{d}\bar{x} f_{\bar{X}_h-\bar{X}_0}(\bar{x})\int_{\mathbf{R}}\mathrm{d}x\, g(x)\left[\int_0^{\bar{x}} f^{(2R+1)}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}(x-t)(\bar{x}-t)^{2R}\mathrm{d}t\right]\\
&= \int_{\mathbf{R}} g(x)r(h,x)\mathrm{d}x,
\end{aligned}$$

where

$$r(h,x) = -\frac{1}{(2R)!}\mathbf{E}\left[\int_0^{\bar{X}_h-\bar{X}_0} f^{(2R+1)}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0}(x-t)(\bar{X}_h-\bar{X}_0-t)^{2R}\mathrm{d}t\right]. \tag{4.21}$$

Plugging these identities in (4.2), we get that, for every test-function g,

$$\mathbf{E}\left[g(\bar{X}_h)\right] = \int_{\mathbf{R}^d} g(x)\left[f_{\bar{X}_0}(x) + f_{\bar{X}_0}(x)\sum_{r=1}^{R}\frac{h^r}{r!}P_r(x) + \widetilde{r}(h,x)\right]\mathrm{d}x,$$

where

$$\widetilde{r}(h,x) = f_{\bar{X}_0}(x)\sum_{r=R+1}^{2R-1}\frac{h^r}{r!}P_r(x) + r(h,x),$$

Hence

$$f_{\bar{X}_h}(x) = f_{\bar{X}_0}(x) + f_{\bar{X}_0}(x)\sum_{r=1}^{R}\frac{h^r}{r!}P_r(x) + \widetilde{r}(h,x). \tag{4.22}$$

The continuity of the function on the right hand side of the above equality will establish the announced expansion, provided we show that $r(h,x) = \mathcal{O}\big(h^{R+\frac{1}{2}}\big)$ uniformly with respect to $x\in\mathbf{R}$. It follows from the boundedness assumption made on $f^{(2R+1)}_{\bar{X}_0\,|\,\bar{X}_h-\bar{X}_0=\bar{x}}(x)$ that

$$\big|r(h,x)\big| \leqslant \frac{1}{(2R)!}\sup_{\bar{x},x\in\mathbf{R}}\Big|f^{(2R+1)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x)\Big|\mathbf{E}\left[\frac{\big|\bar{X}_h-\bar{X}_0\big|^{2R+1}}{2R+1}\right] \leqslant \frac{C_{X,R}}{(2R+1)!}h^{R+\frac{1}{2}}, \tag{4.23}$$

owing to the upper-bound established in (4.11) for $\mathbf{E}\left[|\bar{X}_h - \bar{X}_0|^{2R+1}\right]$.

(b) The claim amounts to integrating Equation (4.22), provided we show that the integrals of $P_r(x)f_{\bar{X}_0}(x)$, for $r = 1, \dots, 2R-1$, are at least semi-convergent and that, for all $b \in \mathbf{R}$, $\int_{-\infty}^{b} r(h,x)dx = \mathcal{O}(h^{R+\frac{1}{2}})$. We first notice that, owing to the assumptions made in (a) and to the upper bound (4.11), we have, for all $a < b \in \mathbf{R}$,

$$\int_a^b \mathbf{E}\left[\int_{\mathbf{R}} \mathbf{1}_{\{t\in I_h\}} \left| f^{(2R+1)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0}(x-t)(\bar{X}_h - \bar{X}_0 - t)^{2R}\right| \mathrm{d}t\right] dx < +\infty,$$

with $I_h := \left[-(\bar{X}_h - \bar{X}_0)^-, (\bar{X}_h - \bar{X}_0)^+\right]$. Hence, owing to Fubini's Theorem, using the definition (4.21) of $r(h,x)$,

$$\int_a^b r(h,x)dx = \\ -\frac{1}{(2R)!}\mathbf{E}\left[\int_0^{\bar{X}_h - \bar{X}_0} \left(f^{(2R)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0}(b-t) - f^{(2R)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0}(a-t)\right)(\bar{X}_h - \bar{X}_0 - t)^{2R}\mathrm{d}t\right].$$

The assumption $\sup_{h\in\mathcal{H},\bar{x},x\in\mathbf{R}} |f^{(2R)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x)| < +\infty$ and again the upper bound (4.11), yield

$$\mathbf{E}\left[\int_{\mathbf{R}} \mathbf{1}_{\{t\in I_h\}} \left| f^{(2R)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0}(a-t)(\bar{X}_h - \bar{X}_0 - t)^{2R}\right| \mathrm{d}t\right] < +\infty.$$

Hence, owing to Lebesgue's Dominated Convergence Theorem and to the assumption $\lim_{x\to-\infty} f^{(2R)}_{\bar{X}_0|\bar{X}_h-\bar{X}_0=\bar{x}}(x) = 0$, we get

$$\int_{-\infty}^{b} r(h,x)dx = -\frac{1}{(2R)!}\mathbf{E}\left[\int_0^{\bar{X}_h - \bar{X}_0} f^{(2R)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0}(b-t)(\bar{X}_h - \bar{X}_0 - t)^{2R}\mathrm{d}t\right] \tag{4.24}$$

and then, likewise (4.23), using the boundedness of $f^{(2R)}_{\bar{X}_0 \mid \bar{X}_h - \bar{X}_0=\bar{x}}(x)$,

$$\int_{-\infty}^{b} r(h,x)dx = \mathcal{O}(h^{R+\frac{1}{2}}). \tag{4.25}$$

Owing to Equation (4.22), if we take $h_1, \dots, h_{2R-1} \in \mathcal{H}$ pairwise distinct, we get, for all $i = 1, \dots, 2R-1$,

$$\sum_{r=1}^{2R-1} h_i^{r-1}\left(\frac{P_r(x)}{r!}f_{\bar{X}_0}(x)\right) = \Xi_i(x), \tag{4.26}$$

with

$$\Xi_i(x) = \frac{f_{\bar{X}_{h_i}}(x) - f_{\bar{X}_0}(x) - r(h_i,x)}{h_i}, \quad i = 1, \dots, 2R-1. \tag{4.27}$$

Hence we get a Vandermonde system, $Vu(x) = \Xi(x)$ with

$$V = V(h_1, \dots, h_{2R-1}) = \begin{pmatrix} 1 & h_1 & h_1^2 & \dots & h_1^{2R-2} \\ 1 & h_2 & h_2^2 & \dots & h_2^{2R-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & h_{2R-1} & h_{2R-1}^2 & \dots & h_{2R-1}^{2R-2} \end{pmatrix},$$

$$u(x) = (u_1(x), \dots, u_{2R-1}(x)) \quad \text{with} \quad u_r(x) = \frac{P_r(x)}{r!}f_{\bar{X}_0}(x)$$

and

$$\Xi(x) = (\Xi_1(x), \dots, \Xi_{2R-1}(x)).$$

We set, for all $j = 1, \dots, 2R-1$,

$$\widetilde{V}_j(x) = \widetilde{V}_j(h_1, \dots, h_{2R-1}, \Xi(x)) = \begin{pmatrix} 1 & \dots & h_1^{j-2} & \Xi_1(x) & h_1^j & \dots & h_1^{2R-2} \\ 1 & \dots & h_2^{j-2} & \Xi_2(x) & h_2^j & \dots & h_2^{2R-2} \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \dots & h_{2R-1}^{j-2} & \Xi_{2R-1}(x) & h_{2R-1}^j & \dots & h_{2R-1}^{2R-2} \end{pmatrix}.$$

By expanding along the j^{th} column, the determinant of $\widetilde{V}_j$ writes

$$\det(\widetilde{V}_j(x)) = \sum_{i=1}^{2R-1} (-1)^{i+j} d_{ij} \Xi_i(x),$$

where $d_{ij} := d_{ij}(h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_{2R}) \in \mathbf{R}$ with

$$d_{ij}(h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_{2R-1}) = \det \begin{pmatrix} 1 & \dots & h_1^{j-2} & h_1^j & \dots & h_1^{2R-2} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & \dots & h_{i-1}^{j-2} & h_{i-1}^j & \dots & h_{i-1}^{2R-2} \\ 1 & \dots & h_{i+1}^{j-2} & h_{i+1}^j & \dots & h_{i+1}^{2R-2} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & \dots & h_{2R-1}^{j-2} & h_{2R-1}^j & \dots & h_{2R-1}^{2R-2} \end{pmatrix}.$$

Hence, owing to Cramer's rule, the solution of the Vandermonde system writes

$$u_j(x) = \frac{\det\left(\widetilde{V}_j(x)\right)}{\det(V)} = \frac{1}{\det(V)} \sum_{i=1}^{2R-1} (-1)^{i+j} d_{ij} \Xi_i(x).$$

Finally, since we saw that for all $i = 1, \dots, 2R-1$, the integral of $\Xi_i(x)$ is semi-convergent, we deduce the semi-convergence of the integral

$$\int_{-\infty}^{b} P_r(x) f_{\bar{X}_0}(x) dx = \frac{r!}{\det(V)} \sum_{i=1}^{2R-1} (-1)^{i+r} d_{ir} \int_{-\infty}^{b} \Xi_i(x) dx,$$

where, owing to the expression (4.27) and to (4.25), $\int_{-\infty}^{b} \Xi_i(x) dx$ is finite, which concludes the proof.

□

4.3.2 Strong convergence rate

We conclude by showing the strong convergence of the nested Monte Carlo estimator. The following Lemma is more or less standard (see for instance [Avi09]).

Lemma 4.3.1. *Let $\bar{X}$ and $\bar{X}'$ be two real valued random variables lying in L^p, $p \geqslant 1$, with densities $f_{\bar{X}}$ and $f_{\bar{X}'}$ respectively. Then, for every $x \in \mathbf{R}$,*

$$\left\| \mathbf{1}_{\{\bar{X} \leqslant x\}} - \mathbf{1}_{\{\bar{X}' \leqslant x\}} \right\|_2^2 \leqslant \left(p^{\frac{p}{p+1}} + p^{\frac{1}{p+1}} \right) \left(\|f_{\bar{X}}\|_{\sup} + \|f_{\bar{X}'}\|_{\sup} \right)^{\frac{p}{p+1}} \|\bar{X} - \bar{X}'\|_p^{\frac{p}{p+1}}. \quad (4.28)$$

Proof. Let $L>0$. Note that

$$\begin{aligned}
\left\|\mathbf{1}_{\{\bar{X}\leqslant x\}}-\mathbf{1}_{\{\bar{X}'\leqslant x\}}\right\|_2^2 &= \mathbf{P}\left(\bar{X}\leqslant x\leqslant \bar{X}'\right)+\mathbf{P}\left(\bar{X}'\leqslant x\leqslant \bar{X}\right)\\
&\leqslant \mathbf{P}\left(\bar{X}\leqslant x,\,\bar{X}'\geqslant x+L\right)+\mathbf{P}\left(\bar{X}\leqslant x\leqslant \bar{X}'\leqslant x+L\right)\\
&\quad +\mathbf{P}\left(\bar{X}'\leqslant x,\bar{X}\geqslant x+L\right)+\mathbf{P}\left(\bar{X}'\leqslant x\leqslant \bar{X}\leqslant x+L\right)\\
&\leqslant \mathbf{P}\left(\bar{X}'-\bar{X}\geqslant L\right)+\mathbf{P}\left(\bar{X}-\bar{X}'\geqslant L\right)\\
&\quad +\mathbf{P}(\bar{X}'\in[x,x+L])+\mathbf{P}(\bar{X}\in[x,x+L])\\
&= \mathbf{P}\left(|\bar{X}'-\bar{X}|\geqslant L\right)+\mathbf{P}(\bar{X}\in[x,x+L])+\mathbf{P}(\bar{X}'\in[x,x+L])\\
&\leqslant \frac{\mathbf{E}\left[|\bar{X}'-\bar{X}|^p\right]}{L^p}+L\Big(\|f_{\bar{X}}\|_{\sup}+\|f_{\bar{X}'}\|_{\sup}\Big).
\end{aligned}$$

A straightforward optimization in L yields the announced result. □

The strong convergence is a consequence of the Proposition 4.3.1 combined with the previous Lemma and Lemma 4.1.2.

Proposition 4.3.2. *Assume $X\in L^p$, $p\geqslant 2$. Under the assumptions of Proposition 4.3.1 (a) with $R=0$ and if the density $f_{\bar{X}_0}$ is bounded, then there exists $h_0=\frac{1}{K_0}\in\mathcal{H}\setminus\{0\}$ such that, for every $h,h'\in(0,h_0]$*

$$\left\|\mathbf{1}_{\{\bar{X}_h\leqslant x\}}-\mathbf{1}_{\{\bar{X}_{h'}\leqslant x\}}\right\|_2^2\leqslant C_{X,Y,p}|h-h'|^{\frac{p}{2(p+1)}},$$

where $C_{X,Y,p}=2^{\frac{p}{p+1}}\left(p^{\frac{p}{p+1}}+p^{\frac{1}{p+1}}\right)(\|f_{\bar{X}_0}\|_{\sup}+1)^{\frac{p}{p+1}}\left(4\,B_p^{MZ}\|X\|_p\right)^{\frac{p}{p+1}}$. This means that the strong approximation error assumption holds with $\beta=\frac{p}{2(p+1)}\in(0,\frac{1}{2})$.

Proof. It follows from Proposition 4.3.1 (a) that

$$f_{\bar{X}_h}(x)\leqslant f_{\bar{X}_0}(x)+o(h^{\frac{1}{2}})\quad\text{uniformly with respect to } x\in\mathbf{R}.$$

Consequently, there exists an $h_0=\frac{1}{K_0}\in\mathcal{H}\setminus\{0\}$ such that, for every $h\in(0,h_0]$,

$$\forall\, x\in\mathbf{R},\ f_{\bar{X}_h}(x)\leqslant f_{\bar{X}_0}(x)+1.$$

Plugging the above bound and (4.8) in Inequality (4.28) of Lemma 4.3.1 applied with $\bar{X}=\bar{X}_h$ and $\bar{X}'=\bar{X}_{h'}$ completes the proof. □

Chapitre 5

Randomized Multilevel Estimator

In this Chapter we compare the unbiased randomized Multilevel estimator proposed by Glynn and Rhee in [RG15] with the Multilevel estimators with and without weights. We will build an unbiased Multilevel estimator using the usual family of biased random variables $(Y_h)_{h\in\mathcal{H}}$ and choosing a random level τ instead of the deterministic R in the Multilevel framework.

Like in Section 1.2, Z_j will denote the contribution for the level j to the estimator under consideration. We will first establish a general result involving these Z_j. As for applications, we will consider two frameworks :

1. First we will set
$$Z_j = Y^{(j)}_{\frac{h}{n_j}} - Y^{(j)}_{\frac{h}{n_{j-1}}}, \quad j \geq 2,$$
with $n_j = M^{j-1}$ and $Z_1 = Y_h$. In this setting the Z_j are independent like we did in Section 1.2.

2. A second case of interest will consist in setting
$$Z_j = Y_{\frac{h}{n_j}} - Y_{\frac{h}{n_{j-1}}},$$
where, by contrast, all the $Y_{\frac{h}{n_j}}$, $j \geq 1$, are sampled from the same family $(Y_h)_{h\in\mathcal{H}}$ (typically in a diffusion setting all Y_h are designed from the same Brownian motion), which yields $\sum_{j=1}^{k} Z_j = Y_{\frac{h}{n_k}}$.

In order to be able to write general results, we introduce the random variables
$$S_m := \sum_{j=1}^{m} Z_j, \quad m \geq 1. \tag{5.1}$$

We recall that in both cases the cost of simulation of Z_j reads
$$\mathrm{Cost}(Z_j) = g\left(\mathrm{Cost}\left(Y_{\frac{h}{n_j}}\right), \mathrm{Cost}\left(Y_{\frac{h}{n_{j-1}}}\right)\right) = g\left(\frac{n_j}{h}, \frac{n_{j-1}}{h}\right) = \kappa\frac{n_j}{h} g(1, M^{-1})$$
and that we make the strong error assumption
$$\left\|Y_h - Y_0\right\|_2^2 \leq V_1 h^\beta.$$

5.1 Unbiased randomized estimator

Let $\tau : (\Omega, \mathcal{A}, \mathbf{P}) \to \mathbf{N}^*$ be a discrete random variable independent of $(Z_j)_{j\geq 1}$. We set

$$I_\tau := \sum_{j=1}^{\tau} \frac{Z_j}{\pi_j},$$

where $\pi_j = \mathbf{P}(\tau \geq j) > 0$ for all $j \geq 1$.

I_τ is an unbiased estimator of $I_0 = \mathbf{E}[Y_0]$, as we show in Proposition 5.1.1. In practice the estimator of interest is designed

$$I_\tau^N := \frac{1}{N} \sum_{k=1}^{N} I_{\tau_k}, \tag{5.2}$$

with

$$I_{\tau_k} = \sum_{j=1}^{\tau_k} \frac{Z_j^{(k)}}{\pi_j},$$

where $(\tau_k)_{k\geq 1}$ are *i.i.d.* with a geometric distribution of parameter $p^* \in (0,1)$ and are independent of the *i.i.d.* sequences $\left((Z_j^{(k)})_{j\geq 1}\right)_{k\geq 1}$. The details about the choice of the optimal parameter p^* are given in the next Section. Though our formalism is borrowed from [RG15], our methods of proof slightly differ from the one developed in this reference.

Proposition 5.1.1. *Assume*

$$\sum_{j\geq 1} \left\|Z_j\right\|_1 < +\infty. \tag{5.3}$$

Then there exists a random variable S_∞ such that $\mathbf{E}[|S_\infty|] < +\infty$, $S_m \to S_\infty$ a.s. and in L^1 as $m \to +\infty$ and

$$\mathbf{E}[I_\tau] = \mathbf{E}[Y_0] = \mathbf{E}[S_\infty].$$

Proof. It is clear by absolute convergence from (5.3) that $S_m \to S_\infty$ *a.s.* and in L^1 as $m \to +\infty$ and $\mathbf{E}[S_\infty] = \mathbf{E}[Y_0]$. We may write

$$I_\tau = \sum_{j\geq 1} \frac{Z_j}{\pi_j} \mathbf{1}_{\{\tau \geq j\}}.$$

Moreover, owing to the independence of τ and $(Z_j)_{j\geq 1}$, we get

$$\sum_{j\geq 1} \frac{\left\|Z_j \mathbf{1}_{\{\tau\geq j\}}\right\|_1}{\pi_j} = \sum_{j\geq 1} \frac{\left\|Z_j\right\|_1 \mathbf{P}(\tau \geq j)}{\pi_j} = \sum_{j\geq 1} \left\|Z_j\right\|_1 < +\infty.$$

Lebesgue's dominated convergence theorem for series yields

$$\mathbf{E}[I_\tau] = \sum_{j\geq 1} \frac{\mathbf{E}\left[Z_j \mathbf{1}_{\{\tau\geq j\}}\right]}{\pi_j} = \sum_{j\geq 1} \mathbf{E}[Z_j] = \mathbf{E}[Y_0].$$

□

The variance of this unbiased estimator is described in the following Proposition.

Proposition 5.1.2. *Assume*

$$\sum_{j\geq 1} \frac{\|Z_j\|_2}{\sqrt{\pi_j}} < +\infty. \tag{5.4}$$

Then there exists a random variable S_∞ such that $\mathbf{E}\left[|S_\infty|^2\right] < +\infty$ and

$$(i) \quad \lim_{m\to\infty} \frac{\|S_\infty - S_m\|_2}{\sqrt{\pi_m}} = 0.$$

$$(ii) \quad \mathrm{Var}(I_\tau) = \mathrm{Var}(S_\infty) + \sum_{j\geq 1} \frac{\|S_\infty - S_j\|_2^2}{\pi_j \pi_{j+1}} \mathbf{P}(\tau = j).$$

Proof. (i) Assumption (5.4) implies that $\sum_{j\geq 1} \|Z_j\|_1 \leq \sum_{j\geq 1} \|Z_j\|_2 < +\infty$, hence S_∞ exists *a.s.* and $\mathbf{E}\left[|S_\infty|^2\right] < +\infty$. Owing to Minkowski's Inequality, to assumption (5.4) and using that π_j is non increasing, we get

$$\frac{\|S_\infty - S_m\|_2}{\sqrt{\pi_m}} \leq \frac{\sum_{j\geq m+1} \|Z_j\|_2}{\sqrt{\pi_m}} \leq \sum_{j\geq m+1} \frac{\|Z_j\|_2}{\sqrt{\pi_j}} \to 0, \quad \text{as} \quad m \to +\infty.$$

(ii) We set

$$I_m := \sum_{j=1}^{m} \frac{Z_j \mathbf{1}_{\{\tau\geq j\}}}{\pi_j}.$$

The assumption (5.4) yields

$$\sum_{j\geq 1} \frac{\|Z_j \mathbf{1}_{\{\tau\geq j\}}\|_2}{\pi_j} = \sum_{j\geq 1} \frac{\|Z_j\|_2 \|\mathbf{1}_{\{\tau\geq j\}}\|_2}{\pi_j} = \sum_{j\geq 1} \frac{\|Z_j\|_2}{\sqrt{\pi_j}} < +\infty,$$

hence $I_m \to I_\tau$ in L^2 as $m \to +\infty$, and

$$\mathbf{E}\left[I_m^2\right] \to \mathbf{E}\left[I_\tau^2\right] \quad \text{as} \quad m \to +\infty. \tag{5.5}$$

Then we derive

$$\begin{aligned}
\mathbf{E}\left[I_m^2\right] &= \sum_{j=1}^{m} \frac{\mathbf{E}\left[Z_j^2\right]}{\pi_j^2}\pi_j + 2 \sum_{1\leq j<\ell\leq m} \frac{\mathbf{E}\left[Z_j Z_\ell\right]}{\pi_j \pi_\ell} \mathbf{P}(\tau \geq j \vee \ell) \\
&= \sum_{j=1}^{m} \frac{\mathbf{E}\left[Z_j^2 + 2Z_j \sum_{\ell=j+1}^{m} Z_\ell\right]}{\pi_j} \\
&= \sum_{j=1}^{m} \frac{\mathbf{E}\left[Z_j + \sum_{\ell=j+1}^{m} Z_\ell\right]^2 - \mathbf{E}\left[\sum_{\ell=j+1}^{m} Z_\ell\right]^2}{\pi_j} \\
&= \sum_{j=1}^{m} \frac{\|S_m - S_{j-1}\|_2^2 - \|S_m - S_j\|_2^2}{\pi_j},
\end{aligned}$$

where we used in the second line that $\mathbf{P}(\tau \geq j \vee \ell) = \mathbf{P}(\tau \geq \ell) = \pi_\ell$, since $j < \ell$. Now,

$$\begin{aligned}\left\|S_m - S_{j-1}\right\|_2^2 - \left\|S_m - S_j\right\|_2^2 &= 2\left\langle Z_j, S_m - \frac{S_j + S_{j-1}}{2}\right\rangle \\ &= 2\left\langle Z_j, S_\infty - \frac{S_j + S_{j-1}}{2}\right\rangle + 2\left\langle Z_j, S_m - S_\infty\right\rangle \\ &= \left\|S_\infty - S_{j-1}\right\|_2^2 - \left\|S_\infty - S_j\right\|_2^2 + 2\left\langle Z_j, S_m - S_\infty\right\rangle.\end{aligned}$$

Consequently,

$$\mathbf{E}\left[I_m^2\right] = \sum_{j=1}^m \frac{\left\|S_\infty - S_{j-1}\right\|_2^2 - \left\|S_\infty - S_j\right\|_2^2}{\pi_j} + 2\sum_{j=1}^m \frac{\left\langle Z_j, S_m - S_\infty\right\rangle}{\pi_j}.$$

One has, by Schwarz's Inequality and the fact that $\sqrt{\pi_j} \geq \sqrt{\pi_m}$ for $j \leq m$,

$$\sum_{j=1}^m \frac{\left\langle Z_j, S_m - S_\infty\right\rangle}{\pi_j} \leq \sum_{j=1}^m \frac{\left\|Z_j\right\|_2}{\sqrt{\pi_j}} \frac{\left\|S_m - S_\infty\right\|_2}{\sqrt{\pi_m}}.$$

Owing to assumption (5.4) and to (i)

$$\lim_{m\to+\infty} \sum_{j=1}^m \frac{\left\langle Z_j, S_m - S_\infty\right\rangle}{\pi_j} = 0,$$

hence

$$\mathbf{E}\left[I_\tau^2\right] = \sum_{j\geq 1} \frac{\left\|S_\infty - S_{j-1}\right\|_2^2 - \left\|S_\infty - S_j\right\|_2^2}{\pi_j}.$$

As a last step, an Abel transform shows that

$$\begin{aligned}\sum_{j\geq 1} \frac{\left\|S_\infty - S_{j-1}\right\|_2^2 - \left\|S_\infty - S_j\right\|_2^2}{\pi_j} &= \frac{\left\|S_\infty\right\|_2^2}{\pi_1} - \frac{\left\|S_\infty - S_m\right\|_2^2}{\pi_m} + \sum_{j=1}^{m-1} \left\|S_\infty - S_j\right\|_2^2 \left(\frac{1}{\pi_{j+1}} - \frac{1}{\pi_j}\right) \\ &= \frac{\left\|S_\infty\right\|_2^2}{\pi_1} - \frac{\left\|S_\infty - S_m\right\|_2^2}{\pi_m} + \sum_{j=1}^{m-1} \frac{\left\|S_\infty - S_j\right\|_2^2}{\pi_j \pi_{j+1}} \mathbf{P}(\tau = j).\end{aligned}$$

Now $\pi_1 = 1$ and (i) yield

$$\mathbf{E}\left[I_\tau\right]^2 = \mathbf{E}\left[S_\infty^2\right] + \sum_{j\geq 1} \frac{\left\|S_\infty - S_j\right\|_2^2}{\pi_j \pi_{j+1}} \mathbf{P}(\tau = j).$$

Combining this with $\mathbf{E}\left[I_\tau\right] = \mathbf{E}\left[Y_0\right] = \mathbf{E}\left[S_\infty\right]$, we finally get the expression for the variance

$$\mathrm{Var}(I_\tau) = \mathrm{Var}(S_\infty) + \sum_{j\geq 1} \frac{\left\|S_\infty - S_j\right\|_2^2}{\pi_j \pi_{j+1}} \mathbf{P}(\tau = j),$$

which completes the proof. □

The analysis of the variance is then reduced to the analysis of the quantity $\left\|S_\infty - S_j\right\|_2^2$. We use S_j^C to refer to the second correlated case and S_j^I to refer to the first independent case. S_∞^C and S_∞^I will denote their corresponding limits (in L^2 and *a.s.*). We start by the second setting.

2. In this fully correlated setting, we have

$$Z_1 = Y_h \quad \text{and} \quad Z_j = Y_{\frac{h}{n_j}} - Y_{\frac{h}{n_{j-1}}}, \quad j \geq 2,$$

so that

$$S_j^C = Y_{\frac{h}{n_j}} \quad \text{and} \quad S_\infty^C = \lim_{m \to +\infty} S_m^C = Y_0$$

in L^2 and *a.s.*. Consequently

$$V_j^C := \left\|S_\infty^C - S_j^C\right\|_2^2 = \left\|Y_0 - Y_{\frac{h}{n_j}}\right\|_2^2 \leq V_1 \left(\frac{h}{n_j}\right)^\beta. \tag{5.6}$$

1. In the independent setting, like for Multilevel estimators MLMC and ML2R, we have that

$$Z_1 = Y_h^{(1)} \quad \text{and} \quad Z_j = Y_{\frac{h}{n_j}}^{(j)} - Y_{\frac{h}{n_{j-1}}}^{(j)}, \quad j \geq 2,$$

are independent, so that S_∞^I is no longer equal to Y_0 (though $\mathbf{E}\left[S_\infty^I\right] = \mathbf{E}\left[Y_0\right]$). The bias-variance decomposition then yields, owing to the independence of the $(Z_\ell)_{\ell \geq 1}$,

$$\begin{aligned} V_j^I &:= \left\|S_\infty^I - S_j^I\right\|_2^2 \\ &= \mathrm{Var}\left(\sum_{i \geq j+1} Z_i\right) + \left(\sum_{i \geq j+1} \mathbf{E}\left[Z_i\right]\right)^2 \\ &= \sum_{i \geq j+1} \mathrm{Var}(Z_i) + \left(\mathbf{E}\left[Y_0\right] - \mathbf{E}\left[Y_{\frac{h}{n_j}}\right]\right)^2 \\ &= \sum_{i \geq j+1} \left\|Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right\|_2^2 + \left(\mathbf{E}\left[Y_0 - Y_{\frac{h}{n_j}}\right]\right)^2 - \sum_{i \geq j+1} \left(\mathbf{E}\left[Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right]\right)^2. \end{aligned} \tag{5.7}$$

Combining (5.6) and (5.7) yields

$$V_j^C - V_j^I = \mathrm{Var}\left(Y_0 - Y_{\frac{h}{n_j}}\right) - \sum_{i \geq j+1} \mathrm{Var}\left(Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right).$$

It is difficult to compare a priori these two quantities V_j^C and V_j^I in full generality. However, if we have the guess that in Setting 2 the random variables Z_j are positively correlated, we see that $V_j^C \geq V_j^I$. Indeed

$$\begin{aligned} V_j^C - V_j^I &= \mathrm{Var}\left(Y_0 - Y_{\frac{h}{n_j}}\right) - \sum_{i \geq j+1} \mathrm{Var}\left(Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right) \\ &= 2 \sum_{i > k \geq j+1} \mathrm{cov}\left(Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}, Y_{\frac{h}{n_k}} - Y_{\frac{h}{n_{k-1}}}\right). \end{aligned}$$

For the variance to be bounded, we need the sum $\sum_{j\geq 1} \frac{\left\|S_\infty - S_j\right\|_2^2}{\pi_j \pi_{j+1}} \mathbf{P}(\tau = j)$ to converge. For this purpose we will use the strong error hypothesis (SE_β) and choose the law of τ in a wise way.

5.2 Geometric distribution for the level

We assume that τ follows a geometric distribution with parameter p, *i.e.* $\mathbf{P}(\tau = j) = (1-p)^{j-1}p$ and $\pi_j = (1-p)^{j-1}$. Hence

$$\frac{\mathbf{P}(\tau = j)}{\pi_j \pi_{j+1}} = p(1-p)^{-j}.$$

We start by the second setting and observe that, owing to the strong error assumption,

$$\mathrm{Var}(I_\tau) \leq \mathrm{Var}(Y_0) + V_1 h^\beta \frac{p}{1-p} \sum_{j\geq 1} \Big((1-p)M^\beta\Big)^{-(j-1)}. \tag{5.8}$$

In order to get the convergence of the series, we need $M^\beta(1-p) > 1$ and get

$$\mathrm{Var}(I_\tau) \leq \mathrm{Var}(Y_0) + V_1 h^\beta \frac{p}{1-p} \frac{(1-p)M^\beta}{(1-p)M^\beta - 1}.$$

We recall that the cost of the levels is given by the formulas

$$\mathrm{Cost}(Z_1) = \frac{\kappa}{h} \leq \frac{\kappa}{h} g(1, M^{-1})$$

and

$$\mathrm{Cost}(Z_j) = \frac{\kappa}{h} \frac{M^{j-1}}{h} g(1, M^{-1}), \quad j \geq 2,$$

hence the cost of the random variable I_τ reads

$$\mathrm{Cost}(I_\tau) \leq \frac{\kappa}{h} g(1, M^{-1}) \sum_{j=1}^{\tau} M^{j-1} = \frac{\kappa}{h} g(1, M^{-1}) \frac{M^\tau - 1}{M-1}. \tag{5.9}$$

Under the sign of expectation formula (5.9) yields

$$\mathbf{E}\left[\mathrm{Cost}(I_\tau)\right] \leq \frac{\kappa}{h} \frac{g(1, M^{-1})}{M-1} \left(Mp \sum_{j\geq 1} \Big(M(1-p)\Big)^{j-1} - 1 \right). \tag{5.10}$$

If we make the additional assumption $M(1-p) < 1$, the series in (5.10) converges and we get

$$\mathbf{E}\left[\mathrm{Cost}(I_\tau)\right] \leq \frac{\kappa}{h} \frac{g(1, M^{-1})}{M-1} \left(Mp \frac{1}{1 - M(1-p)} - 1 \right) = \frac{\kappa}{h} g(1, M^{-1}) \frac{1}{1 - M(1-p)}.$$

Finally, we get the following bounds for p :

$$\frac{1}{M^\beta} < 1 - p < \frac{1}{M}. \tag{5.11}$$

An important remark is that, for these bounds to hold, we will need $\beta > 1$, hence the scope of the randomized Multilevel estimator is limited by this constraint and this estimator cannot be used when $\beta \in (0, 1]$.

The optimal value for the parameter p will be given by the minimization of the effort, as we did for the optimal parameters in the deterministic Multilevel framework.

Owing to the formulas (5.8) and (5.10), we write

$$\mathrm{Var}(I_\tau) \leq \mathrm{Var}(Y_0) + V_1 h^\beta \frac{p}{1-p} \sum_{j\geq 1} \frac{M^{-\beta(j-1)}}{\pi_j}$$

and

$$\mathbf{E}\left[\mathrm{Cost}(I_\tau)\right] \leq \frac{\kappa}{h} \frac{g(1, M^{-1})}{M-1} \left(Mp \sum_{j\geq 1} M^{j-1}\pi_j - 1 \right).$$

We search for p minimizing the effort

$$\phi(I_\tau) = \mathrm{Var}(I_\tau)\mathbf{E}\left[\mathrm{Cost}(I_\tau)\right].$$

This amounts to solving the sub-optimal problem

$$\min_{1-\frac{1}{M}<p<1-\frac{1}{M^\beta}} \left[\phi^*(I_\tau) := \left(V_1 h^\beta \frac{p}{1-p} \sum_{j\geq 1} \frac{M^{-\beta(j-1)}}{\pi_j} \right) \left(\frac{\kappa}{h} \frac{g(1, M^{-1})}{M-1} Mp \sum_{j\geq 1} M^{j-1}\pi_j \right) \right].$$

We ignore the coefficient $\frac{p^2}{1-p}$ and search for the p which minimizes

$$\left(\sum_{j\geq 1} \frac{M^{-\beta(j-1)}}{\pi_j} \right) \left(\sum_{j\geq 1} M^{j-1}\pi_j \right),$$

which, owing to equality case in Cauchy-Schwarz's Inequality, gives

$$\pi_j = (1-p)^{j-1} = \lambda M^{-\frac{\beta+1}{2}(j-1)},$$

for some constant $\lambda > 0$. Since $\pi_1 = \mathbf{P}(\tau \geq 1) = 1$, hence $\lambda = 1$ and we finally get the optimal value for the parameter of the geometric distribution of τ

$$p^* = 1 - M^{-\frac{\beta+1}{2}}. \tag{5.12}$$

Now that we built an unbiased estimator with minimal effort, we can make a Monte Carlo estimator of size N as described in (5.2)

$$I_\tau^N = \frac{1}{N} \sum_{k=1}^{N} I_{\tau_k},$$

where $(I_{\tau_k})_{k\geq 1}$ are independent copies of I_τ. If we fix the constraint on the squared error

$$\left\| I_\tau^N - I_0 \right\|_2 = \varepsilon,$$

we get an optimal size

$$N^* = \mathrm{Var}(I_\tau)\varepsilon^{-2}. \tag{5.13}$$

The cost of the Randomized Multilevel estimator (RML) is then computed with the formula

$$\mathbf{E}\left[\mathrm{Cost}(I_\tau^{N^*})\right] = N^*\mathbf{E}\left[\mathrm{Cost}(I_\tau)\right] = \mathrm{Var}(I_\tau)\varepsilon^{-2}\frac{1+M^{-1}}{1-M(1-p)}.$$

The variance in the first setting being more complex, we will take the same values for p^* and N^* as optimal parameters for the Randomized Multilevel estimator in the independent case.

5.3 Numerical results

We compared the performances of the Randomized Multilevel estimator in its independent and dependent version, on a Call in a Black Scholes model in one dimension, with $X_0 = 100$, $K = 80$, $T = 1$, $\sigma = 0.4$, using a Milstein scheme, in order to guarantee $\beta > 1$.

The optimal parameters are given by formulas (5.12) and (5.13), and the variance $\mathrm{Var}(I_\tau)$ needed to set the optimal size of the estimator in the formula (5.13) is estimated as structural parameter, making a standard Monte Carlo estimator of size 10^6 on I_τ and setting

$$\widehat{\mathrm{Var}}\left(\frac{1}{N}\sum_{n=1}^{N} I_{\tau_n}\right) = \left(\sum_{n=1}^{N}(I_{\tau_n})^2 - \frac{1}{N}(\sum_{n=1}^{N} I_{\tau_n})^2\right)/(N-1).$$

As we did in Chapter 3, in order to check that we respect the theoretical L^2–error, we make a standard Monte Carlo estimator of size $L = 100$ on I_τ^N and we display the empirical L^2–error which we compute with the formula (3.25). The true value is $I_0 = 29.4987$.

We started testing the Randomized Multilevel dependent estimator described in the second setting. In Figure 5.1 we show the results with $M = 2$. The Randomized Multilevel estimator behaves as an unbiased estimator, but the multiplicative constant in front of the cost seems to be bigger than for MLMC and ML2R.

In order to test the sensitivity to the root M, we made M vary and obtained Figure 5.2. $M = 2$ seems not to be the optimal choice for M. An optimal choice for this case seems to be close to $M = 6$.

Hence we fixed $M = 6$ and obtained the results in Figure 5.3, where we can observe the improvement of choosing $M = 6$ with respect to $M = 2$.

To test the robustness of the optimal parameter p, we made it vary between the bounds $\frac{1}{M^\beta} < p < \frac{1}{M}$ given in (5.11) and got Figure 5.4. This Figure confirms that the optimal computed $p^* = 1 - M^{-\frac{\beta+1}{2}} = 0.931959$ is a good choice, the better performances being reached by $p = 0.92, 0.93, 0.94$.

As a second step, we tested the Randomized Multilevel independent estimator described in the first setting and, as we suspected, we obtained that it performs better than the Randomized Multilevel dependent estimator. In Figure 5.5 we show the two Randomized Multilevel estimators, with $M = 6$. At the same time, the Randomized Multilevel estimator in the independent setting does not perform better than ML2R and MLMC, as we can see in Figures 5.6 ($\log_2$ scale on the $x-$axis and $\log_{10}$ scale on the $y-$axis) and 5.7 ($\log_2$ scale on the $x-$axis). In Appendix C.2 we show the Tables of the numerical results for the Multilevel estimators and for the Randomized Multilevel (RML) estimators, dependent and independent case, for $M = 6$ and $p^* = 1 - M^{-\frac{\beta+1}{2}} = 0.931959$. The estimated variances $\mathrm{Var}(I_\tau)$ are given by

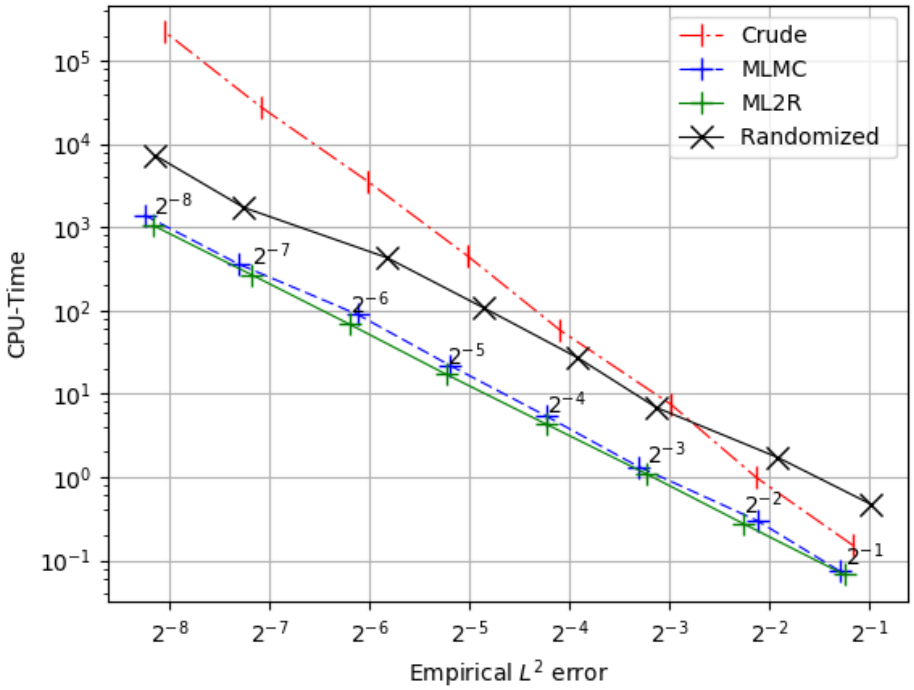


FIGURE 5.1 – In the Randomized Multilevel (RML) estimator $M = 2$ and $p^* = 1 - M^{-\frac{\beta+1}{2}}$ ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

	RML Dependent	RML Independent
$\mathrm{Var}(I_\tau)$	1617.05	1436.61

In conclusion, the randomization of the level in the Randomized Multilevel estimator leads to an unbiased estimator with $\mathrm{Cost}(\varepsilon) \sim C_{RM}\varepsilon^{-2}$, which is the best rate that we can expect, as we saw throughout this work. The multiplicative constant C_{RM}, though, is bigger than the one for MLMC and ML2R estimators, since they return better performances. This constant depends on the variance of I_τ, suggesting that the introduction of randomness via the random variable τ leads to an unbiased framework, but at the same time increases the cost by increasing the variance.

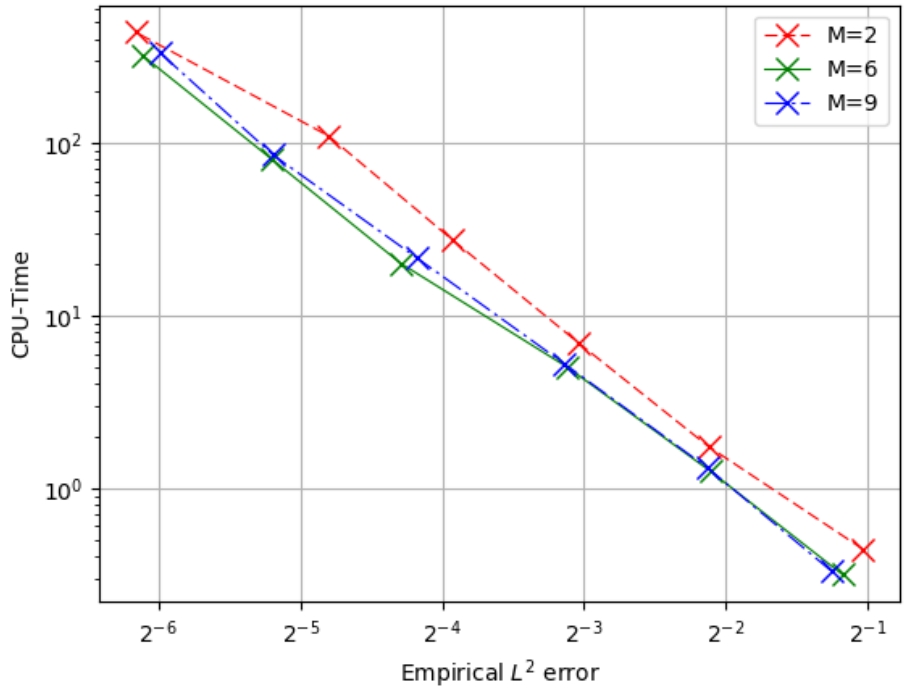


FIGURE 5.2 – In the Randomized Multilevel (RML) estimator M varies and $p^* = 1 - M^{-\frac{\beta+1}{2}}$ ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

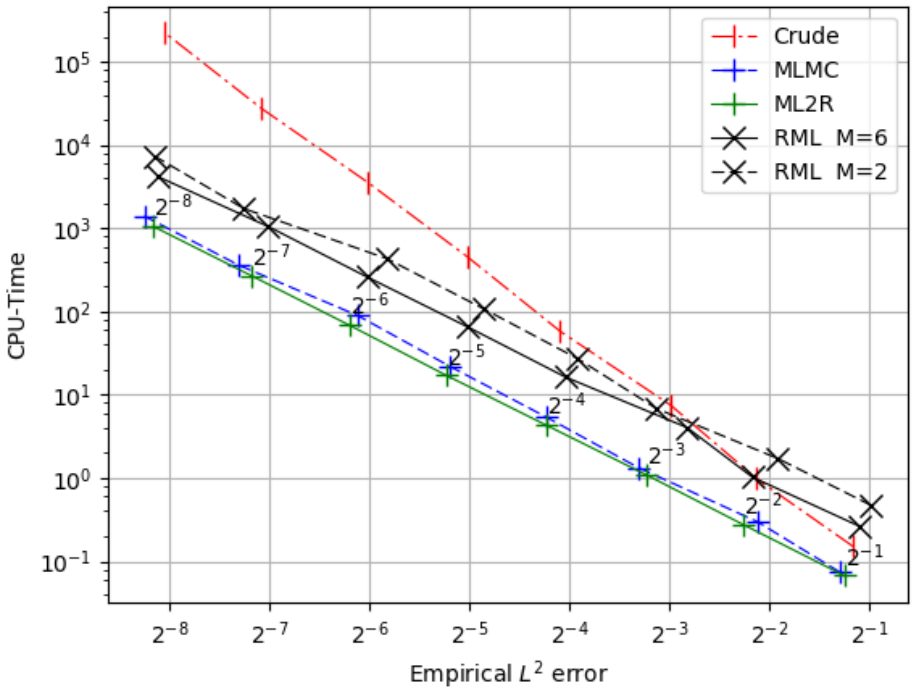


FIGURE 5.3 – In the Randomized Multilevel (RML) estimator $M = 6$ and $M = 2$ ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

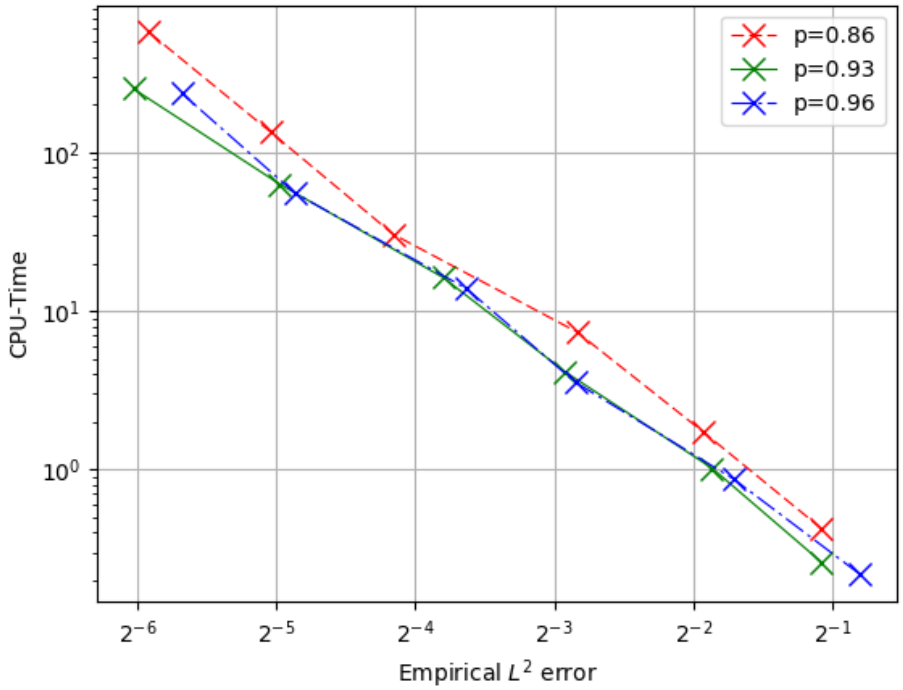


FIGURE 5.4 – In the Randomized Multilevel (RML) estimator $M = 6$ and p varies ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

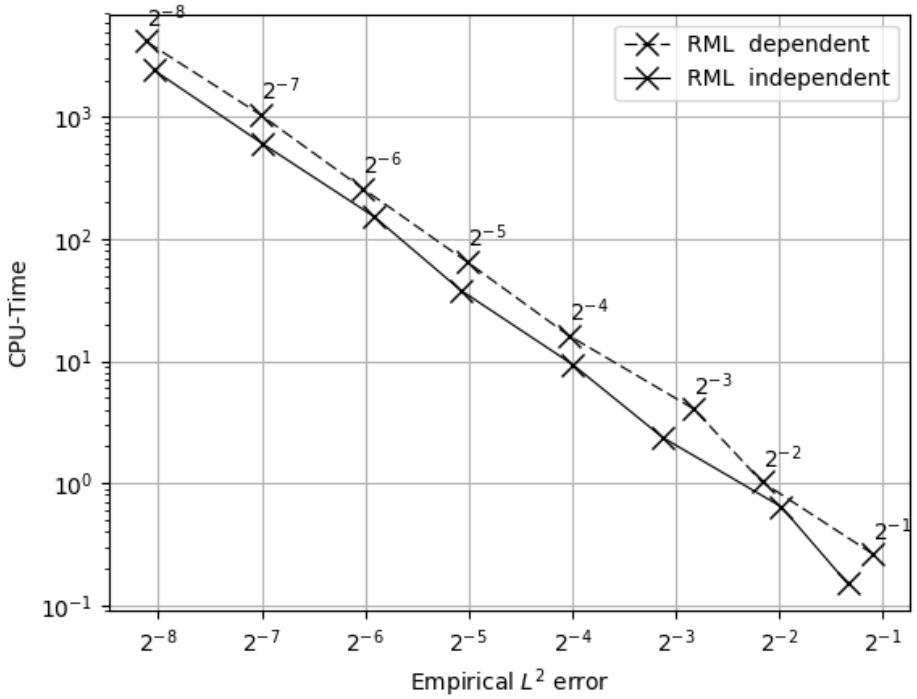


FIGURE 5.5 – Randomized Multilevel (RML) estimator : dependent vs independent version, with $M = 6$ and $p^* = 1 - M^{-\frac{\beta+1}{2}}$ ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

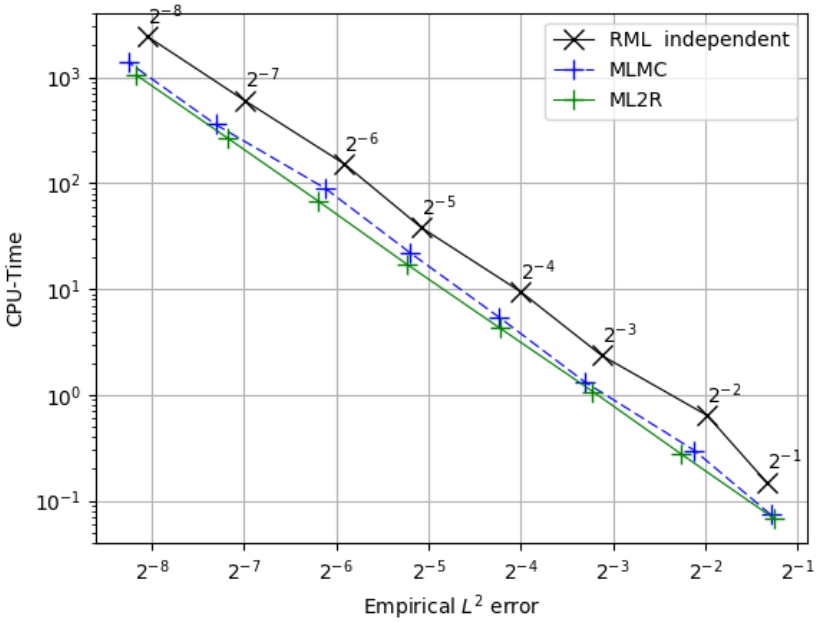


FIGURE 5.6 – Randomized Multilevel (RML) estimator in the independent version, with $M = 6$ and $p^* = 1 - M^{-\frac{\beta+1}{2}}$, vs MLMC and ML2R ($\log_2$ $x-$axis, $\log_{10}$ $y-$axis).

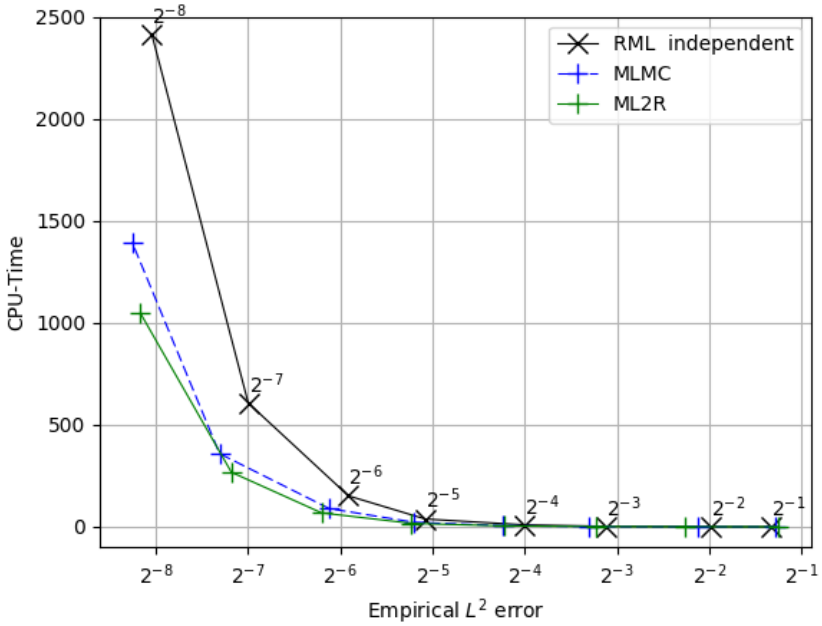


FIGURE 5.7 – Randomized Multilevel (RML) estimator in the independent version, with $M = 6$ and $p^* = 1 - M^{-\frac{\beta+1}{2}}$, vs MLMC and ML2R ($\log_2$ $x-$axis).

Annexe A

Asymptotic of the weights

We focus our attention on the behaviour of $\mathbf{W}_j^R$ when $R \to +\infty$. We recall

$$\mathbf{W}_j^R = \sum_{\ell=j}^{R} a_\ell b_{R-\ell} = \sum_{\ell=0}^{R-j} a_{R-\ell} b_\ell$$

with

$$a_\ell = \frac{1}{\prod_{1\le k\le \ell-1}(1-M^{-k\alpha})}$$

and with the convention $\prod_{k=1}^{0}(1-M^{-k\alpha}) = 1$, and

$$b_\ell = (-1)^\ell \frac{M^{-\frac{\alpha}{2}\ell(\ell+1)}}{\prod_{1\le k\le \ell}(1-M^{-k\alpha})}.$$

For convenience, we set $\mathbf{W}_j^R = 0$, for $j \ge R+1, R \in \mathbf{N}^*$. We first notice that a_ℓ is an increasing and converging sequence and we set

$$\lim_{\ell\to+\infty} a_\ell = a_\infty.$$

The sequence b_ℓ converges to zero and furthermore the series with general term b_ℓ is absolutely converging, since $\sum_{\ell\ge1} M^{-\frac{\alpha}{2}\ell(\ell+1)} < +\infty$. This leads us to set

$$\widetilde{B}_\infty = \sum_{\ell=0}^{+\infty} |b_\ell| < +\infty \quad \text{and} \quad B_\infty = \sum_{\ell=0}^{+\infty} b_\ell < +\infty.$$

Claim (a) of Lemma 2.4.1 is then proved. As a consequence,

$$\forall R \in \mathbf{N}^*, \forall j \in \{1,\dots,R\}, \qquad \left|\mathbf{W}_j^R\right| \le a_\infty \widetilde{B}_\infty, \tag{A.1}$$

which proves claim (b) in Lemma 2.4.1. For the proof of claims (c) and (d), we will need the following

Lemma A.0.1. *Let $\varphi : \mathbf{N}^* \to \mathbf{N}^*$ such that $\varphi(R) \in \{1,\dots,R-1\}$ for every $R \ge 1$, $\varphi(R) \to +\infty$ and $R-\varphi(R) \to +\infty$ as $R \to +\infty$. Then*

$$\lim_{R\to+\infty} \sup_{1\le j\le\varphi(R)} \left|\mathbf{W}_j^R - 1\right| = 0.$$

In particular, $\forall j \in \mathbf{N}^*, \mathbf{W}_j^R \to 1$ as $R \to +\infty$.
However, this convergence is not uniform since $\mathbf{W}_{R-j+1}^R \to a_\infty \sum_{\ell=0}^{j-1} b_\ell$ for every $j \in \mathbf{N}^*$ as $R \to +\infty$.

Proof. We write

$$\left|\mathbf{W}_j^R - a_\infty B_\infty\right| = \left|\sum_{\ell=0}^{R-j} a_{R-\ell} b_\ell - a_\infty \sum_{\ell=0}^{R-j} b_\ell - a_\infty \sum_{\ell \geq R-j+1} b_\ell\right|$$
$$\leq \sum_{\ell=0}^{R-j} (a_\infty - a_{R-\ell}) |b_\ell| + a_\infty \sum_{\ell \geq R-j+1} |b_\ell|$$

First note that

$$\lim_{R\to+\infty} \sup_{j\in\{1,\dots,\varphi(R)\}} \sum_{\ell\geq R-j+1} |b_\ell| \leq \lim_{R\to+\infty} \sum_{\ell\geq R-\varphi(R)+1} |b_\ell| = 0,$$

as $R-\varphi(R)\to+\infty$ and $\sum_{\ell\geq 0}|b_\ell| < +\infty$. On the other hand, for every $j\in\{1,\dots,\varphi(R)\}$,

$$\begin{aligned}
\sum_{\ell=0}^{R-j} (a_\infty - a_{R-\ell}) |b_\ell| &= \sum_{\ell=j}^{R} (a_\infty - a_\ell) |b_{R-\ell}| \\
&= \sum_{\ell=j}^{\varphi(R)} (a_\infty - a_\ell) |b_{R-\ell}| + \sum_{\ell=\varphi(R)+1}^{R} (a_\infty - a_\ell) |b_{R-\ell}| \\
&\leq a_\infty \sum_{\ell=R-\varphi(R)}^{R-j} |b_\ell| + \left(a_\infty - a_{\varphi(R)+1}\right) \sum_{\ell=0}^{R-\varphi(R)-1} |b_\ell| \\
&\leq a_\infty \sum_{\ell=R-\varphi(R)}^{+\infty} |b_\ell| + \left(a_\infty - a_{\varphi(R)+1}\right) \sum_{\ell=0}^{R-\varphi(R)-1} |b_\ell|.
\end{aligned}$$

Consequently, $\sup_{j\in\{1,\dots,R\}} \sum_{\ell=0}^{R-j} (a_\infty - a_{R-\ell}) |b_\ell| \to 0$ as $R\to+\infty$, since $\varphi(R)$ and $R-\varphi(R)\to+\infty$. Finally,

$$\lim_{R\to+\infty} \sup_{1\leq j\leq\varphi(R)} \left|\mathbf{W}_j^R - a_\infty B_\infty\right| = 0.$$

Moreover, by definition $\mathbf{W}_1^R = 1$ for all R, which implies that $B_\infty = \frac{1}{a_\infty}$ and completes the proof. Finally, as $a_j \to a_\infty$,

$$\mathbf{W}_{R-j+1}^R = \sum_{\ell=0}^{j-1} a_{R-\ell} b_l \xrightarrow{R\to+\infty} a_\infty \sum_{\ell=0}^{j-1} b_\ell.$$

□

Proof of Lemma 2.4.1 (c) and (d).

(c) Let us consider the non-negative measure on $\mathbf{N}^*$ defined by $m_\beta(j) = M^{\gamma(j-1)}$, $\gamma < 0$. We notice that it is a finite measure since

$$\sum_{j\geq 1} dm_\beta(j) = \frac{1}{1-M^\gamma}.$$

Since, as we saw in Lemma A.0.1, $\mathbf{W}_j^R \to 1$ as $R \to +\infty$ for every $j \in \mathbf{N}^*$ and $|\,\mathbf{W}_j^R\,| \leq a_\infty \widetilde{B}_\infty$, we derive from Lebesgue's dominated convergence theorem that

$$\lim_{R\to+\infty} \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} = \sum_{j=2}^{+\infty} \lim_{R\to+\infty} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} = \frac{1}{M^{-\gamma}-1}.$$

(d) • If $\gamma < 0$, we consider the non-negative finite measure on $\mathbf{N}^*$ defined by $m'_\beta(j) = M^{\gamma(j-1)} v_j$ since $(v_j)_{j\geqslant 1}$ is a bounded sequence of positive real numbers. As in the previous case (c) we have

$$\lim_{R\to+\infty} \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} v_j = \sum_{j=2}^{+\infty} M^{\gamma(j-1)} v_j.$$

• If $\gamma = 0$, let us consider a sequence $\varphi(R) \in \{1, \ldots, R-1\}$ such that $\frac{\varphi(R)}{R} \to 1$, $R - \varphi(R) \to +\infty$ as $R \to +\infty$ (for example $\varphi(R) = R - \sqrt{R}$). Then we can write

$$\begin{aligned}
&\left| \frac{1}{R} \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| v_j - \frac{1}{R} \sum_{j=2}^{R} v_j \right| \\
&\leq \left[\frac{1}{R} \sum_{j=2}^{\varphi(R)} \left| \left|\mathbf{W}_j^R\right| - 1 \right| + \frac{1}{R} \sum_{j=\varphi(R)+1}^{R} (|\,\mathbf{W}_j^R\,| + 1) \right] \sup_{j\geq 2} v_j \\
&\leq \left[\sup_{2\leq j\leq \varphi(R)} \left|\mathbf{W}_j^R - 1\right| \frac{\varphi(R)}{R} + \left(a_\infty \widetilde{B}_\infty + 1\right) \left(1 - \frac{\varphi(R)}{R}\right) \right] \sup_{j\geq 2} v_j.
\end{aligned}$$

Owing to Lemma A.0.1 $\sup_{2\leq j\leq \varphi(R)} |\,\mathbf{W}_j^R - 1| \to 0$ as $R \to +\infty$. Using furthermore that $\frac{\varphi(R)}{R} \to 1$ as $R \to +\infty$ and that $\lim_{j\to+\infty} v_j = 1$, one concludes by noting that, owing to Césàro's Lemma, $\lim_{R\to+\infty} \frac{1}{R} \sum_{j=2}^{R} v_j = 1$.

• If $\gamma > 0$, first, we notice that

$$\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} \geq \left|\mathbf{W}_R^R\right| M^{\gamma R} = |a_R|\, M^{\gamma R} \to +\infty. \tag{A.2}$$

Let $\eta > 0$. Since $\lim_{j\to+\infty} v_j = 1$, there exists $N_\eta \in \mathbf{N}^*$ such that, for each $j > N_\eta$, $|v_j - 1| < \frac{\eta}{2}$. Owing to Lemma A.0.1 there exists R_η such that, for each $R \geq R_\eta$, $\sup_{2\leq j\leq N_\eta} \left|\mathbf{W}_j^R\right| < 1 + \eta$. Then,

$$
\begin{aligned}
&\left| \frac{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} v_j}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} - 1 \right| \\
&\leq \frac{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} |v_j - 1|}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} \\
&\leq \frac{\sum_{j=2}^{N_\eta} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} |v_j - 1|}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} + \frac{\eta}{2} \frac{\sum_{j=N_\eta+1}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} \\
&\leq \frac{\max_{2\leq j\leq N_\eta} M^{\gamma(j-1)} |v_j - 1| \, N_\eta \sup_{2\leq j\leq N_\eta} \left|\mathbf{W}_j^R\right|}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} + \frac{\eta}{2} \\
&\leq \frac{f(N_\eta)(1+\eta)}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} + \frac{\eta}{2}
\end{aligned}
$$

where $f(N) = \max_{2\leq j\leq N} M^{\gamma(j-1)} |v_j - 1| \, N$ does not depend on R. Thanks to (A.2), there exists $R'_\eta > 0$ such that, for each $R \geq \max(R_\eta, R'_\eta)$, $\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} > \frac{2f(N_\eta)(1+\eta)}{\eta}$, which proves that

$$
\lim_{R\to+\infty} \frac{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} v_j}{\sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)}} = 1.
$$

This leads to analyze

$$
\frac{1}{M^{\gamma R}} \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{\gamma(j-1)} = \sum_{j=2}^{R} \left|\mathbf{W}_j^R\right| M^{-\gamma(R-j+1)} = \sum_{j=1}^{R-1} \left|\mathbf{W}_{R-j+1}^R\right| M^{-\gamma j}.
$$

Using that $\left|\mathbf{W}_j^R\right| \leq a_\infty \widetilde{B}_\infty$ for $j \in \{1, \ldots, R\}$ and Lemma A.0.1 one derives from Lebesgue's dominated convergence theorem that

$$
\sum_{j=1}^{R-1} \left|\mathbf{W}_{R-j+1}^R\right| M^{-\gamma j} \xrightarrow{R\to+\infty} a_\infty \sum_{j\geq 1} \left| \sum_{\ell=0}^{j-1} b_\ell \right| M^{-\gamma j} < +\infty
$$

since $\left|\sum_{\ell=0}^{j-1} b_\ell\right| \leq \sum_{\ell=0}^{j-1} |b_\ell| \leq \widetilde{B}_\infty$.

□

Annexe B

Closed formulas in Clark-Cameron model

We consider a Clark-Cameron model

$$\begin{cases} dU_t = S_t dW_{1,t} \\ dS_t = \mu dt + dW_{2,t} \end{cases} \tag{B.1}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent Brownian motions, with initial conditions $U_0 = S_0 = 0$, which is equivalent to

$$\begin{cases} U_t = \int_0^t (\mu s + W_{2,s}) dW_{1,s} \\ S_t = \mu t + W_{2,t} \end{cases} \tag{B.2}$$

B.1 Norm payoff

We aim at computing $\mathbf{E}\left[U_T^2 + S_T^2\right]$. First of all $\mathbf{E}\left[S_T^2\right] = \mathbf{E}\left[(\mu T + W_{2,T})^2\right] = \mu^2 T^2 + T$. The value of $\mathbf{E}\left[U_T^2\right]$ follows from the independence of the two Brownian motions, indeed

$$\mathbf{E}\left[U_T^2\right] = \mathbf{E}\left[\left(\int_0^T \mu s dW_{1,s}\right)^2\right] + 2\mathbf{E}\left[\int_0^T \mu s dW_{1,s} \int_0^T W_{2,s} dW_{1,s}\right] + \mathbf{E}\left[\left(\int_0^T W_{2,s} dW_{1,s}\right)^2\right].$$

Owing to the independence of the two Brownian motions

$$\mathbf{E}\left[\int_0^T s dW_{1,s} \int_0^T W_{2,s} dW_{1,s}\right] = \mathbf{E}\left[\phi(W_{2,s})\right]$$

with

$$\phi((w_s)_{0\le s\le T}) = \mathbf{E}\left[\int_0^T s dW_{1,s} \int_0^T w_s dW_{1,s}\right] = \int_0^T s w_s ds,$$

hence

$$\mathbf{E}\left[\int_0^T s dW_{1,s} \int_0^T W_{2,s} dW_{1,s}\right] = \mathbf{E}\left[\int_0^T s W_{2,s} ds\right] = 0.$$

This yields

$$\mathbf{E}\left[U_T^2\right] = \mu^2 \frac{T^3}{3} + \frac{T^2}{2}.$$

Adding a constant volatility to the first term, $dU_t = \sigma S_t dW_{1,t}$ yields

$$\mathbf{E}\left[U_T^2\right] = \sigma^2\mu^2 \frac{T^3}{3} + \sigma^2 \frac{T^2}{2}.$$

Finally we obtained

$$\mathbf{E}\left[U_T^2 + S_T^2\right] = T + \left(\mu^2 + \frac{\sigma^2}{2}\right) T^2 + \sigma^2\mu^2 \frac{T^3}{3}.$$

B.2 Cosinus payoff

We aim at computing $\mathbf{E}\left[\cos(\lambda U_t)\right]$. Conditionnaly to the process $(W_{2,s})_{0\leq s\leq t}$, U_t has a centered gaussian distribution with stochastic variance

$$\int_0^t (\mu s + W_{2,s})^2 ds.$$

Let us compute $\mathbf{E}\left[e^{-i\lambda U_t}\right]$. Owing to the independence of W_1 and W_2, we write

$$\begin{aligned}
\mathbf{E}\left[e^{-i\lambda U_t}\right] &= \mathbf{E}\left[e^{-i\lambda\int_0^t(\mu s+W_{2,s})dW_{1,s}}\right] = \mathbf{E}\left[\mathbf{E}\left[e^{-i\lambda\int_0^t(\mu s+W_{2,s})dW_{1,s}}|(W_{2,s})_{0\leq s\leq t}\right]\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[e^{-i\lambda\int_0^t(\mu s+w_s)dW_{1,s}}\right]_{(w_s)_{0\leq s\leq t}=(W_{2,s})_{0\leq s\leq t}}\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[e^{-\frac{\lambda^2}{2}\int_0^t(\mu s+w_s)^2 ds}\right]_{(w_s)_{0\leq s\leq t}=(W_{2,s})_{0\leq s\leq t}}\right] \\
&= \mathbf{E}\left[e^{-\frac{\lambda^2}{2}\int_0^t(\mu s+W_{2,s})^2 ds}\right].
\end{aligned}$$

We apply Girsanov's Theorem and we make a change of probability $\mathbf{P}^*$ such that under $\mathbf{P}^*$ $B_s = \mu s + W_{2,s}$ is a Brownian motion, *i.e.* $\frac{d\mathbf{P}}{d\mathbf{P}^*} = e^{\mu B_t - \frac{1}{2}\mu^2 t}$, which yields

$$\begin{aligned}
\mathbf{E}\left[e^{-i\lambda U_t}\right] &= \mathbf{E}^*\left[e^{\mu B_t-\frac{1}{2}\mu^2 t}e^{-\frac{\lambda^2}{2}\int_0^t(B_s)^2 ds}\right] = \mathbf{E}^*\left[e^{\mu B_t-\frac{1}{2}\mu^2 t}\mathbf{E}^*\left[e^{-\frac{\lambda^2}{2}\int_0^t(B_s)^2 ds}|B_t\right]\right] \\
&= e^{-\frac{1}{2}\mu^2 t}\mathbf{E}^*\left[e^{\mu B_t}\mathbf{E}^*\left[e^{-\frac{\lambda^2}{2}\int_0^t(B_s)^2 ds}|B_t^2\right]\right]. \qquad \text{(B.3)}
\end{aligned}$$

We recall formula p.427 in [PY82]

$$\mathbf{E}_x\left(e^{-\frac{b^2}{2}\int_0^t B_s^2 ds}|B_t^2 = y\right) = \left(\frac{bt}{\sinh bt}e^{\frac{y}{2t}(1-bt\coth bt)}\right)\frac{I_{-\frac{1}{2}}\left(\frac{zb}{\sinh bt}\right)}{I_{-\frac{1}{2}}\left(\frac{z}{t}\right)}$$

with $z=\sqrt{xy}$ and where I_ν is the Bessel function which, for $\nu=-1/2$, reads $I_{-\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi}}\frac{\cosh z}{\sqrt{z}}$. Here $x=0$, hence

$$\mathbf{E}^*\left[e^{-\frac{\lambda^2}{2}\int_0^t(B_s)^2 ds}|B_t^2 = y\right] = \sqrt{\frac{\lambda t}{\sinh\lambda t}}e^{\frac{y}{2t}(1-\lambda t\coth\lambda t)}.$$

Plugging this in the formula (B.3) yields

$$\mathbf{E}\left[e^{-i\lambda U_t}\right] = \sqrt{\frac{\lambda t}{\sinh\lambda t}}e^{-\frac{1}{2}\mu^2 t}\mathbf{E}^*\left[e^{\mu B_t+\frac{B_t^2}{2t}(1-\lambda t\coth\lambda t)}\right].$$

Since $B_t\sim\mathcal{N}(0,t)$,

$$\begin{aligned}
\mathbf{E}^*\left[e^{\mu B_t+\frac{B_t^2}{2t}(1-\lambda t\coth\lambda t)}\right] &= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi t}}e^{\mu x+\frac{x^2}{2t}(1-\lambda t\coth\lambda t)}e^{-\frac{x^2}{2t}}\mathrm{d}x \\
&= \frac{1}{\sqrt{t}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(x^2\lambda\coth\lambda t - 2\mu x\right)\right).
\end{aligned}$$

We set $a = \sqrt{\lambda \coth \lambda t}$ and $b = \mu/a$ and we get

$$\mathbf{E}^* \left[e^{\mu B_t + \frac{B_t^2}{2t}(1-\lambda t \coth \lambda t)} \right] = \frac{1}{\sqrt{t}} e^{\frac{b^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(ax-b)^2}{2} \right) \mathrm{d}x = \frac{1}{\sqrt{t}} e^{\frac{b^2}{2}} \frac{1}{a}.$$

Hence

$$\mathbf{E}\left[e^{-i\lambda U_t} \right] = \sqrt{\frac{\lambda t}{\sinh \lambda t}} \frac{1}{\sqrt{t}} \frac{1}{\sqrt{\lambda \coth \lambda t}} e^{-\frac{1}{2}\mu^2 t + \frac{b^2}{2}} = \frac{1}{\sqrt{\cosh \lambda t}} e^{-\frac{\mu^2 t}{2}\left(1 - \frac{\tanh \lambda t}{\lambda t}\right)}.$$

Since the right term of the above equality has no imaginary part, we obtained

$$\mathbf{E}\left[\cos(\lambda U_t) \right] = \frac{1}{\sqrt{\cosh \lambda t}} e^{-\frac{\mu^2 t}{2}\left(1 - \frac{\tanh \lambda t}{\lambda t}\right)}. \tag{B.4}$$

We notice that these computations are shortened when there is no drift term

$$\begin{cases} dU_t = S_t dW_{1,t} \\ dS_t = dW_{2,t} \end{cases}$$

Indeed, owing to the Cameron-Martin formula (see [RY94] p.445),

$$\mathbf{E}\left[e^{-\lambda \int_0^1 (W_{2,s})^2 ds} \right] = \left(\cosh \sqrt{2\lambda} \right)^{-\frac{1}{2}}$$

and the scaling property of the Brownian motion yields

$$\mathbf{E}\left[e^{-i\lambda U_t} \right] = \mathbf{E}\left[e^{-\frac{\lambda^2}{2} \int_0^t (W_{2,s})^2 ds} \right] = \mathbf{E}\left[e^{-\frac{\lambda^2 t^2}{2} \int_0^1 (W_{2,s})^2 ds} \right] = (\cosh \lambda t)^{-\frac{1}{2}}.$$

Adding a constant volatility to the Levy's area, $dU_t = \sigma S_t dW_{1,t}$ is equivalent to take $\lambda = \lambda\sigma$ in formula (B.4).

Annexe C

Tables

C.1 Antithetic schemes

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.63e-01	1.84e-01	-7.70e-02	6.40e-02	2	3	2	1.06e+03	4.65e+03
3	1.25e-01	1.13e-01	8.23e-01	-3.62e-02	1.17e-02	2	5	2	4.14e+03	2.19e+04
4	6.25e-02	5.63e-02	4.03e+00	-4.74e-03	3.18e-03	2	9	2	1.76e+04	1.19e+05
5	3.12e-02	2.58e-02	2.20e+01	-2.75e-04	6.73e-04	3	3	2	9.77e+04	6.60e+05
6	1.56e-02	1.22e-02	8.68e+01	-1.96e-04	1.51e-04	3	3	2	3.91e+05	2.64e+06
7	7.81e-03	7.77e-03	3.92e+02	-1.48e-03	5.88e-05	3	4	2	1.57e+06	1.21e+07
8	3.91e-03	3.26e-03	1.76e+03	-7.54e-04	1.02e-05	3	5	2	6.42e+06	5.50e+07
9	1.95e-03	1.70e-03	7.94e+03	-4.28e-04	2.72e-06	3	6	2	2.64e+07	2.48e+08
10	9.77e-04	8.62e-04	3.56e+04	3.75e-05	7.50e-07	3	7	2	1.09e+08	1.10e+09

Table C.1 – ML2R estimator with Euler scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.12e-01	1.29e-01	3.40e-05	4.52e-02	4	2	1	9.27e+02	3.65e+03
3	1.25e-01	1.13e-01	5.17e-01	-1.40e-02	1.28e-02	4	2	1	3.71e+03	1.46e+04
4	6.25e-02	5.75e-02	2.06e+00	2.65e-03	3.33e-03	4	2	1	1.48e+04	5.82e+04
5	3.12e-02	2.74e-02	8.24e+00	-2.40e-03	7.54e-04	5	2	1	6.09e+04	2.52e+05
6	1.56e-02	1.35e-02	3.32e+01	1.44e-03	1.82e-04	5	2	1	2.44e+05	1.01e+06
7	7.81e-03	7.12e-03	1.29e+02	-3.68e-04	5.10e-05	5	2	1	9.74e+05	4.04e+06
8	3.91e-03	3.41e-03	5.15e+02	-1.65e-04	1.17e-05	5	2	1	3.90e+06	1.61e+07
9	1.95e-03	1.73e-03	2.14e+03	-1.47e-04	3.02e-06	6	2	1	1.59e+07	6.79e+07
10	9.77e-04	8.68e-04	8.33e+03	-8.89e-05	7.52e-07	6	2	1	6.34e+07	2.72e+08

Table C.2 – ML2R estimator with Giles-Szpruch scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.58e-01	1.48e-01	1.64e-02	6.68e-02	3	2	1	1.32e+03	4.14e+03
3	1.25e-01	1.22e-01	5.91e-01	6.87e-03	1.51e-02	3	2	1	5.26e+03	1.66e+04
4	6.25e-02	6.16e-02	2.32e+00	1.47e-04	3.83e-03	3	2	1	2.10e+04	6.62e+04
5	3.12e-02	2.90e-02	1.11e+01	-2.55e-03	8.44e-04	4	2	1	9.43e+04	3.42e+05
6	1.56e-02	1.49e-02	4.39e+01	4.67e-04	2.24e-04	4	2	1	3.77e+05	1.37e+06
7	7.81e-03	7.39e-03	1.71e+02	8.94e-04	5.43e-05	4	2	1	1.51e+06	5.47e+06
8	3.91e-03	4.11e-03	6.84e+02	-6.24e-04	1.67e-05	4	2	1	6.04e+06	2.19e+07
9	1.95e-03	1.85e-03	2.98e+03	-2.24e-04	3.41e-06	5	2	1	2.58e+07	1.02e+08
10	9.77e-04	1.01e-03	1.18e+04	-8.01e-05	1.02e-06	5	2	1	1.03e+08	4.06e+08

TABLE C.3 – ML2R estimator with Ninomiya-Victoir scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.02e-01	3.99e-01	5.71e-02	3.81e-02	3	4	2	1.66e+03	1.13e+04
3	1.25e-01	9.77e-02	1.92e+00	2.15e-02	9.18e-03	3	5	2	7.10e+03	5.51e+04
4	6.25e-02	5.72e-02	9.64e+00	2.53e-02	2.66e-03	3	7	2	3.14e+04	2.97e+05
5	3.12e-02	2.50e-02	4.62e+01	1.31e-02	4.55e-04	3	9	2	1.36e+05	1.48e+06
6	1.56e-02	1.24e-02	2.56e+02	3.67e-03	1.43e-04	4	6	2	6.86e+05	8.34e+06
7	7.81e-03	5.96e-03	1.17e+03	2.40e-03	3.01e-05	4	7	2	2.88e+06	3.85e+07
8	3.91e-03	3.38e-03	5.66e+03	1.38e-03	9.63e-06	5	5	2	1.34e+07	1.88e+08
9	1.95e-03	1.78e-03	2.65e+04	1.01e-03	2.18e-06	5	6	2	5.69e+07	8.96e+08
10	9.77e-04	8.72e-04	1.38e+05	4.20e-04	5.90e-07	6	5	2	2.63e+08	4.51e+09

TABLE C.4 – MLMC estimator with Euler scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.33e-01	1.41e-01	1.34e-01	3.66e-02	5	2	1	1.12e+03	4.11e+03
3	1.25e-01	1.06e-01	6.38e-01	6.37e-02	7.27e-03	6	2	1	4.87e+03	1.91e+04
4	6.25e-02	5.14e-02	2.71e+00	2.58e-02	1.99e-03	7	2	1	2.05e+04	8.46e+04
5	3.12e-02	3.08e-02	1.04e+01	1.59e-02	7.04e-04	8	2	1	8.51e+04	3.62e+05
6	1.56e-02	1.50e-02	4.52e+01	7.87e-03	1.64e-04	9	2	1	3.49e+05	1.52e+06
7	7.81e-03	7.20e-03	1.88e+02	4.01e-03	3.61e-05	10	2	1	1.42e+06	6.28e+06
8	3.91e-03	3.75e-03	7.57e+02	1.93e-03	1.04e-05	11	2	1	5.74e+06	2.57e+07
9	1.95e-03	1.78e-03	3.07e+03	9.39e-04	2.32e-06	12	2	1	2.31e+07	1.04e+08
10	9.77e-04	9.34e-04	1.24e+04	5.89e-04	5.31e-07	13	2	1	9.31e+07	4.23e+08

TABLE C.5 – MLMC estimator with Giles-Szpruch scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.25e-01	1.11e-01	1.07e-02	5.11e-02	3	2	1	8.97e+02	3.27e+03
3	1.25e-01	1.21e-01	4.43e-01	4.85e-02	1.25e-02	3	2	1	3.59e+03	1.31e+04
4	6.25e-02	6.65e-02	2.03e+00	1.73e-02	4.16e-03	4	2	1	1.61e+04	6.26e+04
5	3.12e-02	3.04e-02	8.22e+00	1.21e-02	7.89e-04	4	2	1	6.43e+04	2.51e+05
6	1.56e-02	1.33e-02	3.47e+01	3.54e-03	1.66e-04	5	2	1	2.73e+05	1.12e+06
7	7.81e-03	8.14e-03	1.38e+02	3.74e-03	5.28e-05	5	2	1	1.09e+06	4.47e+06
8	3.91e-03	3.21e-03	5.35e+02	-1.26e-04	1.04e-05	6	2	1	4.54e+06	1.92e+07
9	1.95e-03	1.91e-03	2.16e+03	4.60e-04	3.48e-06	6	2	1	1.82e+07	7.69e+07
10	9.77e-04	9.35e-04	9.02e+03	2.00e-04	8.43e-07	7	2	1	7.45e+07	3.23e+08

TABLE C.6 – MLMC estimator with Giles-Szpruch Ninomiya-Victoir scheme. Payoff function $f(u, s) = 10\cos(u)$.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
2	2.50e-01	2.19e-01	1.20e-01	2.64e-02	4.79e-02	3	2	1	1.23e+03	3.47e+03
3	1.25e-01	1.07e-01	4.78e-01	2.34e-02	1.10e-02	3	2	1	4.90e+03	1.38e+04
4	6.25e-02	5.76e-02	2.96e+00	6.83e-03	3.31e-03	4	2	1	2.40e+04	8.03e+04
5	3.12e-02	2.87e-02	9.79e+00	1.20e-02	6.84e-04	4	2	1	9.59e+04	3.21e+05
6	1.56e-02	1.49e-02	4.56e+01	4.10e-03	2.08e-04	5	2	1	4.32e+05	1.61e+06
7	7.81e-03	6.71e-03	1.81e+02	1.95e-03	4.16e-05	5	2	1	1.73e+06	6.45e+06
8	3.91e-03	3.78e-03	8.11e+02	9.25e-04	1.36e-05	6	2	1	7.45e+06	2.99e+07
9	1.95e-03	1.83e-03	3.24e+03	6.35e-04	2.99e-06	6	2	1	2.98e+07	1.20e+08
10	9.77e-04	8.32e-04	1.40e+04	1.36e-04	6.81e-07	7	2	1	1.25e+08	5.27e+08

TABLE C.7 – MLMC estimator with Ninomiya-Victoir scheme. Payoff function $f(u, s) = 10\cos(u)$.

C.2 Randomized Multilevel estimators

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	N	Cost
1	5.00e-01	4.70e-01	2.62e-01	4.10e-02	2.22e-01	6.47e+03	1.28e+04
2	2.50e-01	2.23e-01	1.02e+00	-2.92e-02	4.94e-02	2.59e+04	5.10e+04
3	1.25e-01	1.41e-01	4.06e+00	3.25e-03	2.02e-02	1.03e+05	2.04e+05
4	6.25e-02	6.10e-02	1.62e+01	7.50e-03	3.70e-03	4.14e+05	8.16e+05
5	3.12e-02	3.09e-02	6.53e+01	5.02e-04	9.61e-04	1.66e+06	3.26e+06
6	1.56e-02	1.53e-02	2.60e+02	-1.54e-03	2.35e-04	6.62e+06	1.31e+07
7	7.81e-03	7.75e-03	1.04e+03	6.21e-05	6.06e-05	2.65e+07	5.22e+07
8	3.91e-03	3.59e-03	4.18e+03	7.25e-05	1.30e-05	1.06e+08	2.09e+08

TABLE C.8 – Randomized estimator with Milstein scheme. Dependent case. Call Black-Scholes.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	N	Cost
1	5.00e-01	3.96e-01	1.50e-01	-1.10e-02	1.58e-01	5.75e+03	1.13e+04
2	2.50e-01	2.54e-01	6.40e-01	5.04e-03	6.50e-02	2.30e+04	4.53e+04
3	1.25e-01	1.15e-01	2.36e+00	-3.35e-03	1.33e-02	9.19e+04	1.81e+05
4	6.25e-02	6.26e-02	9.42e+00	1.02e-02	3.85e-03	3.68e+05	7.25e+05
5	3.12e-02	2.96e-02	3.81e+01	-2.66e-03	8.80e-04	1.47e+06	2.90e+06
6	1.56e-02	1.66e-02	1.52e+02	-1.85e-04	2.80e-04	5.88e+06	1.16e+07
7	7.81e-03	7.84e-03	6.07e+02	-4.24e-04	6.19e-05	2.35e+07	4.64e+07
8	3.91e-03	3.80e-03	2.41e+03	9.72e-04	1.36e-05	9.41e+07	1.86e+08

TABLE C.9 – Randomized estimator with Milstein scheme. Independent case. Call Black-Scholes.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	5.00e-01	4.22e-01	6.73e-02	-1.24e-01	1.63e-01	2	5	1	9.14e+03	1.10e+04
2	2.50e-01	2.09e-01	2.75e-01	1.07e-02	4.37e-02	2	9	1	3.66e+04	4.53e+04
3	1.25e-01	1.07e-01	1.08e+00	1.45e-02	1.13e-02	3	5	1	1.43e+05	1.82e+05
4	6.25e-02	5.34e-02	4.29e+00	-2.50e-03	2.84e-03	3	5	1	5.72e+05	7.27e+05
5	3.12e-02	2.66e-02	1.71e+01	-2.54e-04	7.10e-04	3	5	1	2.29e+06	2.91e+06
6	1.56e-02	1.36e-02	6.84e+01	1.64e-03	1.82e-04	3	6	1	9.11e+06	1.16e+07
7	7.81e-03	6.90e-03	2.67e+02	1.06e-04	4.77e-05	4	6	1	3.58e+07	4.65e+07
8	3.91e-03	3.49e-03	1.05e+03	5.05e-04	1.19e-05	4	6	1	1.43e+08	1.86e+08

TABLE C.10 – ML2R estimator with Milstein scheme. Call Black-Scholes.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	5.00e-01	4.10e-01	7.38e-02	-1.74e-01	1.37e-01	2	4	1	1.04e+04	1.19e+04
2	2.50e-01	2.30e-01	2.99e-01	-1.33e-01	3.54e-02	2	7	1	4.22e+04	5.02e+04
3	1.25e-01	1.01e-01	1.33e+00	-4.08e-02	8.54e-03	3	4	1	1.81e+05	2.23e+05
4	6.25e-02	5.33e-02	5.39e+00	-2.41e-02	2.26e-03	3	6	1	7.21e+05	9.08e+05
5	3.12e-02	2.73e-02	2.19e+01	-1.27e-02	5.85e-04	3	8	1	2.90e+06	3.72e+06
6	1.56e-02	1.44e-02	8.93e+01	-8.26e-03	1.39e-04	4	5	1	1.19e+07	1.53e+07
7	7.81e-03	6.32e-03	3.58e+02	-1.89e-03	3.64e-05	4	7	1	4.74e+07	6.17e+07
8	3.91e-03	3.27e-03	1.40e+03	-1.22e-03	9.23e-06	4	8	1	1.90e+08	2.49e+08

Table C.11 – MLMC estimator with Milstein scheme. Call Black-Scholes.

k	$\varepsilon = 2^{-k}$	L^2 error	Time(s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	5.00e-01	4.48e-01	1.51e-01	-1.61e-01	1.75e-01	1	1	4	7.55e+03	3.02e+04
2	2.50e-01	2.30e-01	9.77e-01	-8.86e-02	4.48e-02	1	1	7	2.97e+04	2.08e+05
3	1.25e-01	1.27e-01	7.44e+00	-6.79e-02	1.14e-02	1	1	14	1.18e+05	1.65e+06
4	6.25e-02	5.87e-02	5.77e+01	-2.37e-02	2.89e-03	1	1	28	4.69e+05	1.31e+07
5	3.12e-02	3.08e-02	4.54e+02	-1.50e-02	7.26e-04	1	1	56	1.87e+06	1.05e+08
6	1.56e-02	1.54e-02	3.56e+03	-7.49e-03	1.82e-04	1	1	111	7.47e+06	8.29e+08
7	7.81e-03	7.36e-03	2.77e+04	-2.93e-03	4.55e-05	1	1	222	2.98e+07	6.63e+09
8	3.91e-03	3.78e-03	2.20e+05	-1.71e-03	1.14e-05	1	1	444	1.19e+08	5.30e+10

Table C.12 – CRUDEMC estimator with Milstein scheme. Call Black-Scholes.

Bibliographie

[AGJC16] A. Al Gerbi, B. Jourdain et E. Clément. « Ninomiya-Victoir scheme : strong convergence, antithetic version and application to multilevel estimators ». In : *Monte Carlo Methods Appl.* 22.3 (2016), p. 197–228.

[Avi09] R. Avikainen. « On irregular functionals of SDEs and the Euler scheme ». In : *Finance and Stochastics* 13.3 (2009), p. 381–401.

[BAK15] M. Ben Alaya et A. Kebaier. « Central limit theorem for the multilevel Monte Carlo Euler method ». In : *Ann. Appl. Probab.* 25.1 (fév. 2015), p. 211–234.

[BHR15] K. Bujok, B. Hambly et C. Reisinger. « Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives ». In : *Methodology and Computing in Applied Probability* 17.3 (2015), p. 579–604.

[Bil99] P. Billingsley. *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics : Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, p. x+277.

[BT96a] V. Bally et D. Talay. « The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function ». In : *Probab. Theory Related Fields* 104.1 (1996), p. 43–60.

[BT96b] V. Bally et D. Talay. « The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density ». In : *Monte Carlo Methods Appl.* 2.2 (1996), p. 93–128.

[CL12] N. Chen et Y. Liu. « Estimating expectations of functionals of conditional expected via multilevel nested simulation ». In : *Presentation at conference on Monte Carlo and Quasi-Monte Carlo Methods, Sydney*. 2012.

[Com74] L. Comtet. *Advanced combinatorics*. enlarged. The art of finite and infinite expansions. D. Reidel Publishing Co., Dordrecht, 1974, p. xi+343.

[DG95] D. Duffie et P. Glynn. « Efficient Monte Carlo simulation of security prices ». In : *The Annals of Applied Probability* (1995), p. 897–905.

[DL09] L. Devineau et S. Loisel. « Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique Solvabilité II ». In : *Bulletin Français d'Actuariat, Institut des Actuaires, 10 (17)* (2009), p. 188–221.

[DR15] K. Debrabant et A. Rössler. « On the acceleration of the multi-level Monte Carlo method ». In : *J. Appl. Probab.* 52.2 (2015), p. 307–322.

[Gil08] M. Giles. « Multilevel Monte Carlo path simulation ». In : *Oper. Res.* 56.3 (2008), p. 607–617.

[Gil15] M. B. Giles. « Multilevel Monte Carlo methods ». In : *Acta Numer.* 24 (2015), p. 259–328.

[GS14] M. B. Giles et L. Szpruch. « Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation ». In : *Ann. Appl. Probab.* 24.4 (2014), p. 1585–1620.

[HA12] A. L. Haji Ali. « Pedestrian Flow in the Mean Field Limit ». Mém.de mast. King Abdullah University of Science and Technology, Thuwal, 2012.

[HH80] P. Hall et C. C. Heyde. *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980, p. xii+308.

[JP98] J. Jacod et P. Protter. « Asymptotic error distributions for the Euler method for stochastic differential equations ». In : *Ann. Probab.* 26.1 (jan. 1998), p. 267–307.

[Keb05] A. Kebaier. « Statistical Romberg extrapolation : a new variance reduction method and applications to option pricing ». In : *Ann. Appl. Probab.* 15.4 (2005), p. 2681–2705.

[LP17] V. Lemaire et G. Pagès. « Multilevel Richardson–Romberg extrapolation ». In : *Bernoulli* 23.4A (2017), p. 2643–2692.

[NV08] S. Ninomiya et N. Victoir. « Weak approximation of stochastic differential equations and application to derivative pricing ». In : *Appl. Math. Finance* 15.1-2 (2008), p. 107–121.

[OTc12] K. Oshima, J. Teichmann et D. Velu˘ s˘ cek. « A new extrapolation method for weak approximation schemes with applications ». In : *Ann. Appl. Probab.* 22.3 (2012), p. 1008–1045.

[Pag07] G. Pagès. « Multi-step Richardson-Romberg extrapolation : remarks on variance control and complexity ». In : *Monte Carlo Methods Appl.* 13.1 (2007), p. 37–70.

[PY82] J. Pitman et M. Yor. « A decomposition of Bessel bridges ». In : *Z. Wahrsch. Verw. Gebiete* 59.4 (1982), p. 425–457.

[RG15] C.-H. Rhee et P. W. Glynn. « Unbiased estimation with square root convergence for SDE models ». In : *Oper. Res.* 63.5 (2015), p. 1026–1043.

[RY94] D. Revuz et M. Yor. *Continuous martingales and Brownian motion*. Second. T. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1994, p. xii+560.

[Shi96] A. N. Shiryaev. *Probability*. Second. T. 95. Graduate Texts in Mathematics. Translated from the first (1980) Russian edition by R. P. Boas. New York : Springer-Verlag, 1996, p. xvi+623.

[TT90] D. Talay et L. Tubaro. « Expansion of the global error for numerical schemes solving stochastic differential equations ». In : *Stochastic Anal. Appl.* 8.4 (1990), 483–509 (1991).