Research Article

Daphné Giorgi, Vincent Lemaire and Gilles Pagès*

Limit theorems for weighted and regular Multilevel estimators

DOI: 10.1515/mcma-2017-0102
Received November 17, 2016; accepted January 26, 2017

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 [7] and of its weighted version, the Multilevel Richardson–Romberg estimator (ML2R), introduced in [12]. These two estimators permit to compute a very accurate approximation of $I_0 = \mathbb{E}[Y_0]$ by a Monte Carlo-type estimator when the (non-degenerate) random variable $Y_0 \in L^2(\mathbb{P})$ 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.

Keywords: Multilevel Monte Carlo methods, law of large numbers, Central Limit Theorem, Richardson–Romberg extrapolation, Euler scheme, nested Monte Carlo

MSC 2010: Primary 65C05; secondary 65C30

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 [8] 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 estimator, Multilevel Monte Carlo estimator (MLMC) introduced by Giles in [7] and the Multilevel Richardson–Romberg (weighted) estimator introduced in [12]. 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 [2] in the framework of Euler discretization schemes of Brownian diffusions. We will also deduce an SLLN and a CLT for Multilevel nested Monte Carlo, which are new results to our knowledge. More generally, our result apply to any implementation of Multilevel Monte Carlo methods.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(Y_h)_{h\in\mathcal{H}}$ be a family of real-valued random variables in $L^2(\mathbb{P})$ associated to Y_0, where $\mathcal{H} = \{\frac{\mathbf{h}}{n} : n \geq 1\}$, such that $\lim_{h\to 0}\|Y_h - Y_0\|_2 = 0$. 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

Daphné Giorgi, Vincent Lemaire: Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC Paris 6 (Sorbonne Université), France, e-mail: daphne.giorgi@upmc.fr, vincent.lemaire@upmc.fr
***Corresponding author: Gilles Pagès:** Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC Paris 6 (Sorbonne Université), France, e-mail: gilles.pages@upmc.fr

of $I_0 = \mathbb{E}[Y_0]$ is the standard Monte Carlo estimator, which reads for a fixed h,

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

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 $h_j = \frac{h}{n_j}$ with $n_j = M^{j-1}$, $j = 1, \ldots, R$. If N is the estimator size, we consider an allocation policy $q = (q_1, \ldots, q_R)$ such that, at each level $j = 1, \ldots, R$, we will simulate $N_j = \lceil Nq_j \rceil$ scenarios (see (1) and (2) below). Thus, we consider R independent copies of the family $Y^{(j)} = (Y_h^{(j)})_{h\in\mathcal{H}}$, $j = 1, \ldots, R$, 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 $\pi \in \Pi \subset \mathbb{R}^d$.

A standard Multilevel Monte Carlo (MLMC) estimator, as introduced by Giles in [7], reads

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

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

A Multilevel Richardson–Romberg (ML2R) estimator, as introduced in [12], is a weighted version of (1) which reads

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

with $\pi = (h, R, q)$. The weights $(\mathbf{W}_j^R)_{j=1,\ldots,R}$ are explicitly defined as functions of the weak error rate α (see equation ($\mathrm{WE}_{\alpha,\bar{R}}$) below) and of the refiners n_j, $j = 0, \ldots, R$, in order to kill the successive bias terms in the weak error expansion (see Section 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 $\mathbf{W}_j^R = 1$, $j = 1, \ldots, R$. 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,

$$(\pi(\varepsilon), N(\varepsilon)) = \operatorname*{argmin}_{\|I_\pi^N - I_0\|_2 \leq \varepsilon} \mathrm{Cost}(I_\pi^N). \tag{3}$$

This calibration has been done in [12] 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

$$\exists\, \alpha > 0,\ \bar{R} \geqslant 1,\ (c_r)_{1\leq r\leq \bar{R}}, \quad \mathbb{E}[Y_h] - \mathbb{E}[Y_0] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + h^{\alpha\bar{R}}\eta_{\bar{R}}(h), \quad \lim_{h\to 0}\eta_{\bar{R}}(h) = 0. \tag{$\mathrm{WE}_{\alpha,\bar{R}}$}$$

The second one, called *strong approximation error assumption*, states

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

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

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

From now on, we set $I_\pi^N(\varepsilon) := I_{\pi(\varepsilon)}^{N(\varepsilon)}$, where $\pi(\varepsilon)$ and $N(\varepsilon)$ are closed to solutions of (3) (see [12] for the construction of these parameters and Tables 1 and 2 for the explicit values). As mentioned by Duffie and Glynn in [5], the global cost of the standard Monte Carlo with these *optimal* parameters satisfies

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

where the finite real constant $K(\alpha)$ depends on the structural parameters α, $\mathrm{Var}(Y_0)$, $\mathbf{h}$, 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 [7] and Lemaire and Pagès for ML2R in [12] 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

$$\mathrm{Cost}(I_\pi^N(\varepsilon)) \lesssim K(\alpha,\beta,M)v(\varepsilon), \tag{4}$$

where the finite real constant $K(\alpha,\beta,M)$ is explicit and differs between MLMC and ML2R (see [12] for more details). Denoting v_{MLMC} and v_{ML2R} the dominated function in (4) 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. $v_{\mathrm{MLMC}}(\varepsilon) = v_{\mathrm{ML2R}}(\varepsilon) = \varepsilon^{-2}$. In the case $\beta \leqslant 1$, the ML2R is asymptotically quite better than MLMC since $\lim_{\varepsilon\to 0}\frac{v_{\mathrm{ML2R}}}{v_{\mathrm{MLMC}}} = 0$. More precisely, we have the following scheme.

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

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 $\|I_\pi^N(\varepsilon) - I_0\|_2 \leq \varepsilon$, the convergence in L^2 holds by construction. As a consequence, it is straight forward that, for every sequence $(\varepsilon_k)_{k\geq 1}$ such that $\sum_{k\geq 1}\varepsilon_k^2 < +\infty$,

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

so that

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

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

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

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

$$\|I_\pi^N(\varepsilon) - I_0\|_2^2 = \mu(\varepsilon)^2 + \mathrm{Var}(I_\pi^N(\varepsilon)) \leq \varepsilon^2. \tag{5}$$

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 3.

The paper is organized as follows. In Section 2 we briefly recall the technical background for Multilevel Monte Carlo estimators. In Section 3 we stable our main results: a Strong Law of Large Numbers and a Central Limit Theorem in a quite general framework. Section 4 is devoted to the analysis of the asymptotic behavior 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 5. In Section 6 we apply these results first to the discretization schemes of Brownian diffusions, where we retrieve recent results by Ben Alaya and Kebaier in [2], and secondly to Nested Monte Carlo.

Notations.
- Let $\mathbb{N}^* = \{1, 2, \dots\}$ denote the set of positive integers and $\mathbb{N} = \mathbb{N}^* \cup \{0\}$.
- For every $x \in \mathbb{R}_+ = [0, +\infty)$, $\lceil x \rceil$ denotes the unique $n \in \mathbb{N}^*$ satisfying $n - 1 < x \leqslant n$.
- If $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ 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 $(\varepsilon_n)_{n\in\mathbb{N}}$ 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 Brief background on MLMC and ML2R estimators

We follow [12] and recall briefly the construction of the *optimal* parameters derived from the optimization problem (3). The first step is a stratification procedure allowing us to establish the optimal allocation policy $(q_1, \dots, q_R)$ 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

$$Z_j = \Big(\frac{h}{M^{j-1}}\Big)^{-\frac{\beta}{2}}\big(Y^{(j)}_{\frac{h}{M^{j-1}}} - Y^{(j)}_{\frac{h}{M^{j-2}}}\big) \quad \text{for all } j \geqslant 2 \qquad \text{and} \qquad Z_1 = Y^{(1)}_h,$$

a Multilevel estimator MLMC (1) or ML2R (2) writes

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

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

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

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

$$\begin{cases} q_1^* = \mu^* \sqrt{\frac{\mathrm{Var}(Y_h)}{\mathrm{Cost}(Y_h)}} \\ q_j^* = \mu^* (\frac{h}{M^{j-1}})^{\frac{\beta}{2}} |W_j^R| \sqrt{\frac{\mathrm{Var}(Z_j)}{\mathrm{Cost}(Z_j)}} \end{cases} \qquad \text{with } \mu^* \text{ such that } \sum_{j=1}^{R} q_j = 1, \tag{6}$$

using the Schwarz's inequality (see [12, Theorem 3.6] for a detailed proof). The strong error assumption (SE_β) allows us to upper bound $\mathrm{Var}(Y_h)$ and $\mathrm{Var}(Z_j)$ by $\mathrm{Var}(Y_0)(1 + \theta h^{\beta/2})^2$ with $\theta = \sqrt{\frac{V_1}{\mathrm{Var}(Y_0)}}$ and $V_1(1 + M^{\beta/2})^2$, respectively. On the other hand, we assume that $\mathrm{Cost}(Z_j) = \frac{(1+M^{-1})}{h} M^{j-1}$. Plugging theses estimates in (6), we obtain the *optimal* allocation policy used in this paper and given in Tables 1 and 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 $\mathrm{Var}(Z_j)$ with $V_1 |1 - M^{\beta/2}|^2$ and a new expression for the q_j follows. In the same spirit, the $\mathrm{Cost}(Z_j)$ can be different and hence have an impact on the q_j, see [6] 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 $(\mathrm{WE}_{\alpha,\bar{R}})$ and we obtain

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

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

Table 1. Optimal parameters for the ML2R estimator.

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

Table 2. Optimal parameters for the MLMC estimator.

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

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

The depth parameter $R \geqslant 2$ follows and the choice of N is directly related to the constraint (5).

We report in Tables 1 and 2 the ML2R and MLMC values for $R(\varepsilon)$, $h(\varepsilon)$, $q(\varepsilon) = (q_1(\varepsilon),\dots,q_R(\varepsilon))$, $N(\varepsilon)$ computed in [12] 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 1 and 2:

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

and

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

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.

In what follows, we will shorter these notations by setting

$$R(\varepsilon) = \left\lceil C_R^{(1)} + \sqrt{C_R^{(2)} + \frac{2}{\alpha\log(M)}\log\left(\frac{1}{\varepsilon}\right)}\right\rceil \tag{7}$$

with

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

for ML2R and

$$R(\varepsilon) = \left\lceil C_R^{(1)} + \frac{1}{\alpha\log(M)}\log\left(\frac{1}{\varepsilon}\right)\right\rceil \tag{8}$$

with

$$C_R^{(1)} = 1 + \frac{\log(\lvert c_1\rvert^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \frac{\log(A)}{\alpha\log(M)}$$

for MLMC.

3 Main results

The asymptotic behavior, as ε goes to 0, of the parameters given in Tables 1 and 2 will be exposed in Section 4. We proceed here to the analysis of the asymptotic behavior of the estimator $I_\pi^N(\varepsilon) := I_{\pi(\varepsilon)}^{N(\varepsilon)}$ as $\varepsilon \to 0$.

3.1 Strong Law of Large Numbers

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

Theorem 3.1 (Strong Law of Large Numbers). *Let $p \geq 2$. Assume* $(\mathrm{WE}_{\alpha,\bar{R}})$ *for all $\bar{R} \geq 1$ and $Y_0 \in L^p$. Assume furthermore the following L^p-strong error rate assumption:*

$$\exists\, \beta > 0,\ V_1^{(p)} \geqslant 0, \quad \|Y_h - Y_0\|_p^p = \mathbb{E}[|Y_h - Y_0|^p] \leqslant V_1^{(p)} h^{\frac{\beta}{2}p}, \quad h \in \mathcal{H}. \tag{9}$$

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

$$I_\pi^N(\varepsilon_k) \xrightarrow{\text{a.s.}} I_0 \quad \text{as } k \to +\infty. \tag{10}$$

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

$$Z(h) := \Big(\frac{h}{M}\Big)^{-\frac{\beta}{2}} (Y_{\frac{h}{M}} - Y_h) \quad \text{and} \quad Z_j := Z\Big(\frac{\mathbf{h}}{n_{j-1}}\Big). \tag{11}$$

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

3.2.1 Case $\beta > 1$

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

Theorem 3.2 (Central Limit Theorem, $\beta > 1$). *Assume (SE_β) for $\beta > 1$ and that $(Z(h))_{h\in\mathcal{H}}$ is L^2-uniformly integrable. We set*

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

with

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

Then the following statements hold.

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

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_1^2 + \sigma_2^2) \quad \text{as } \varepsilon \to 0. \tag{12}$$

(b) *MLMC estimator: Assume $(\mathrm{WE}_{\alpha,\bar{R}})$ 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*

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

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, $\operatorname{Var}(Z_j)$, $j \geq 2$, contribute throughout σ_2^2. The ML2R estimator is asymptotically unbiased, whereas the MLMC estimator has an a priori non-vanishing bias term. This gain on the bias for ML2R is balanced by the variance, which is reduced of a factor $\frac{2\alpha}{1+2\alpha}$ for MLMC. The constraint (5) yields $\sigma_1^2 + \sigma_2^2 \leq 1$, which is easy to verify if we recall that

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

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

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

$$\mathbb{E}[Z_j] = \Big(\frac{\mathbf{h}}{n_j}\Big)^{-\frac{\beta}{2}} \mathbb{E}\big[Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_j}}\big] = c_1(1 - M^\alpha)\Big(\frac{\mathbf{h}}{n_j}\Big)^{\alpha-\frac{\beta}{2}} + o\Big(\Big(\frac{\mathbf{h}}{n_j}\Big)^{\alpha-\frac{\beta}{2}}\Big),$$

so that

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

We conclude that

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

Theorem 3.3 (Central Limit Theorem, $0 < \beta \leqslant 1$). *Assume* (SE_β) *for* $\beta \in (0, 1]$. *Assume that* $(Z(h))_{h\in\mathcal{H}}$ *is* L^2*-uniformly integrable and assume furthermore* $\lim_{h\to 0}\|Z(h)\|_2^2 = v_\infty(M, \beta)$. *We set*

$$\sigma^2 = \begin{cases} v_\infty(M, \beta)(1 + M^{\frac{\beta}{2}})^{-2} V_1^{-1} & \text{if } 2\alpha > \beta, \\ (v_\infty(M, \beta) - c_1^2(1 - M^{\frac{\beta}{2}})^2)(1 + M^{\frac{\beta}{2}})^{-2} V_1^{-1} & \text{if } 2\alpha = \beta. \end{cases} \tag{13}$$

Then the following statements hold.

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

$$\frac{I_\pi^N(\varepsilon) - I_0}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \textit{as } \varepsilon \to 0. \tag{14}$$

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

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

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

3.3 Practitioner's corner

In the proof of Theorems 3.2 and 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 4.5 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 $(\varepsilon\sqrt{N(\varepsilon)})^{-1} \to 0$ 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 (7), $\varepsilon\sqrt{N(\varepsilon)} \sim C(\log(1/\varepsilon))^{\frac{1}{4}}$, where C is a positive constant when $\beta = 1$ and $\varepsilon\sqrt{N(\varepsilon)} = o(\varepsilon^{-\eta})$ for all $\eta > 0$ when $\beta < 1$. Hence the convergence rate to 0 of $(\varepsilon\sqrt{N(\varepsilon)})^{-1}$ 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 $(\varepsilon\sqrt{N(\varepsilon)})^{-1}\Sigma_1$ 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 behavior 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 4.5 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.

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 behavior as $\varepsilon \to 0$ of the optimal parameters given in Tables 1 and 2 and of the bias of the estimators and we analyze the weights of the ML2R estimator.

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 [12] 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 [12] and define the polynomial

$$P(R) = \frac{R(R-1)}{2}\log(M) - R\log(K) - \frac{1}{\alpha}\log\Big(\frac{\sqrt{1+4\alpha}}{\varepsilon}\Big), \tag{15}$$

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

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

We notice that

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

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

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

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

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

Let us analyze the denominator

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

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

$$\Big(\frac{1+4\alpha}{1+2\alpha R^*}\Big)^{-\frac{1}{2\alpha R^*}} e^{-\frac{P(R^*)}{R^*}} \leq 1 \quad \text{for all } \varepsilon \in (0, \bar{\varepsilon}), \tag{16}$$

which yields

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

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 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 $\sqrt{\frac{2}{\alpha \log(M)} \log(\frac{1}{\varepsilon})}$ in the ML2R case and $\frac{1}{\alpha \log(M)} \log(\frac{1}{\varepsilon})$ in the MLMC case.

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

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

is the bias of the estimator. This leads us to analyze carefully its asymptotic behavior as $\varepsilon \to 0$. Under assumption $(\mathrm{WE}_{\alpha,\bar{R}})$, 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 [12]:

Proposition 4.2. *The following statements hold.*

(a) *MLMC: Assume* $(\mathrm{WE}_{\alpha,\bar{R}})$ *with* $\bar{R} = 1$.

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

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

(b) *ML2R: Assume* $(\mathrm{WE}_{\alpha,\bar{R}})$ *for all* $\bar{R} \geq 1$.

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

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 $(\mathrm{WE}_{\alpha,\bar{R}})$, whereas the MLMC estimator only needs a first order expansion. As the computations were made under the constraint $\|I_\pi^N - I_0\|_2 \leq \varepsilon$, 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 (15) and (16) we get, for ML2R,

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

and, for MLMC,

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

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

$$\begin{cases} \lim_{\varepsilon\to 0} m(\varepsilon) = 0 & \text{for ML2R,} \\ \frac{M^{-\alpha}}{\sqrt{1+2\alpha}} < \lim_{\varepsilon\to 0} m(\varepsilon) \leq \frac{1}{\sqrt{1+2\alpha}} & \text{for MLMC.} \end{cases} \tag{17}$$

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 (17) still holds. The only uncertain case is when the growth of c_R is faster than polynomial. Then, if $\hat{c}_\infty \geq |c_R|^{\frac{1}{R}}$, $|\frac{\mu(\varepsilon)|}{\varepsilon}$ 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

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

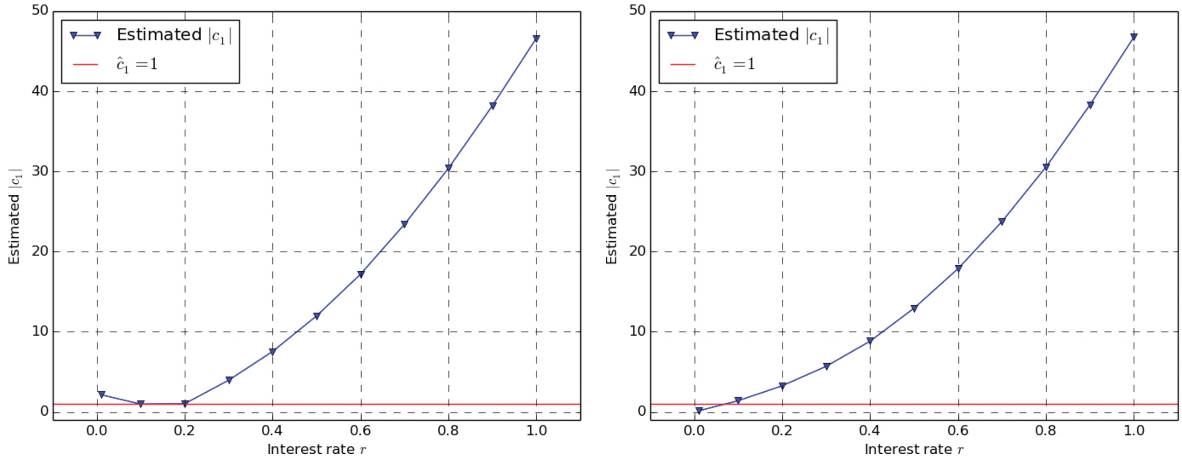
hence $\hat{c}_\infty < 1$ is definitely not a good choice. In conclusion, whenever the growth of c_R is at most polynomial, $\hat{c}_\infty = 1$ remains a good choice. When the growth is faster than polynomial, it is better to overestimate $\hat{c}_\infty$ than to underestimate it. The remarkable fact 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 $\frac{|\mu(\varepsilon)|}{\varepsilon} \leq \frac{1}{\sqrt{1+2\alpha}}$ 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 1a and 1b we show the values of $|c_1|$ estimated with the formula

$$c_1 = \Big(\frac{h}{2} - h\Big)^{-1} (\mathbb{E}[Y_{\frac{h}{2}}] - \mathbb{E}[Y_h])$$

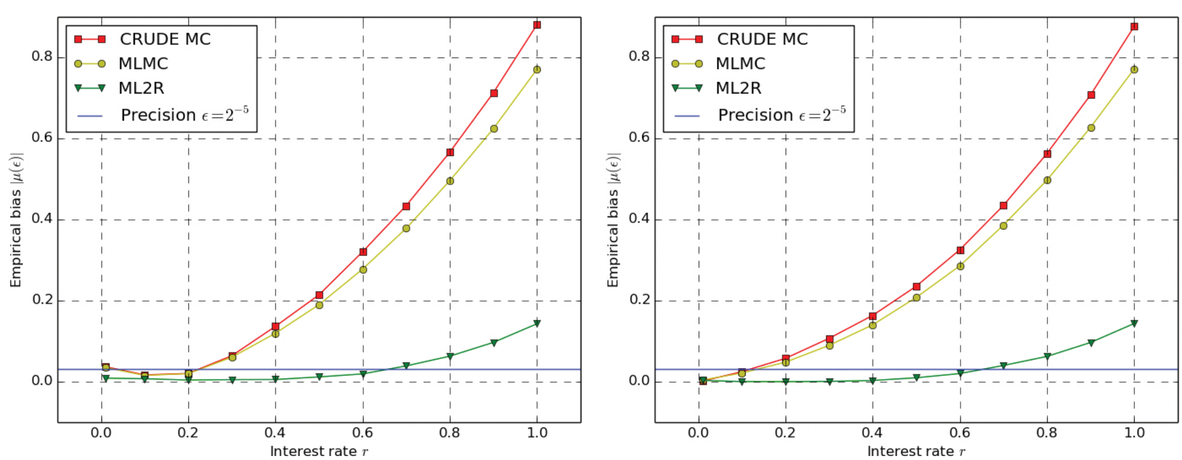
compared to the value plugged in the simulations $\hat{c}_1 = 1$, for a Call option in a Black–Scholes model with $X_0 = 100$, $K = 80$, $T = 1$, $\sigma = 0.4$ and making the interest rate vary as follows: $r = 0.01, 0.1, 0.2, \ldots, 0.9, 1$. We simulated $\mathbb{E}[Y_h]$, with $h = \frac{T}{20}$,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

In Figures 2a and 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 $\tilde{c}$.



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

Figure 1. Estimated $|c_1| = (|\mathbb{E}[Y_h] - \mathbb{E}[Y_{\frac{h}{2}}]|)(h - \frac{h}{2})^{-1}$ when r varies.



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

Figure 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 $\hat{c}_\infty = \hat{c}_1 = 1$.

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 3.2 and 3.3, is to deal with the weights $\mathbf{W}_j^R$ appearing in the ML2R estimator. Moreover, the analysis of the behavior of the weights is necessary when studying the asymptotic of the parameters $q = (q_1, \dots, q_R)$ and N. These weights are devised to kill the coefficients $c_1, \dots, c_R$ in the bias expansion under ($\mathrm{WE}_{\alpha,\bar{R}}$). They are defined as

$$\mathbf{W}_j^R = \sum_{r=j}^{R} \mathbf{w}_r, \quad j = 1, \dots, R, \tag{18}$$

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

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

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 4.3. *Let $\alpha > 0$ and the associated weights $(\mathbf{W}_j^R)_{j=1,\dots,R}$ given in* (18).

(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,*

$$|\mathbf{W}_j^R| \le a_\infty \widetilde{B}_\infty \quad \textit{for all } R \in \mathbb{N}^* \textit{ and all } j \in \{1, \dots, R\}. \tag{19}$$

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

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

(d) *Let $\{v_j\}_{j\ge1}$ be a bounded sequence of positive real numbers. Let $\gamma \in \mathbb{R}$ and assume that* $\lim_{j\to+\infty} v_j = 1$ *when $\gamma \ge 0$. Then the following limits hold:*

$$\sum_{j=2}^{R} |\mathbf{W}_j^R| M^{\gamma(j-1)} v_j \sim \begin{cases} \sum_{j\ge2} M^{\gamma(j-1)} v_j < +\infty & \textit{for} \quad \gamma < 0, \\ R & \textit{for} \quad \gamma = 0, \\ M^{\gamma R} a_\infty \sum_{j\ge1} |\sum_{\ell=0}^{j-1} b_\ell| M^{-\gamma j} & \textit{for} \quad \gamma > 0, \end{cases} \quad \textit{as } R \to +\infty.$$

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

$$q_1(\varepsilon) = \mu^*(\varepsilon)(1 + \theta \mathbf{h}^{\frac{\beta}{2}}) \quad \text{and} \quad q_j(\varepsilon) = \theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta} \mu^*(\varepsilon) |\mathbf{W}_j^{R(\varepsilon)}| M^{-\frac{\beta+1}{2}(j-1)}, \quad j = 2, \dots, R(\varepsilon), \tag{20}$$

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

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

Owing to Lemma 4.3 (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:

$$\frac{1}{\bar{\mu}^*} := 1 + \theta \mathbf{h}^{\frac{\beta}{2}} \le \frac{1}{\mu^*(\varepsilon)} \le 1 + \theta \mathbf{h}^{\frac{\beta}{2}} \left(1 + \underline{C}_{M,\beta} a_\infty \widetilde{B}_\infty \frac{1}{1 - M^{-\frac{\beta+1}{2}}}\right) =: \frac{1}{\underline{\mu}^*}. \tag{21}$$

Remark 4.4. If we set $\mathbf{W}_j^{R(\varepsilon)} = 1$ for all $j = 1, \ldots, R(\varepsilon)$, 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 4.5. *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*

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

- *Case $\beta \leq 1$: We recall the expression of $R(\varepsilon)$ given in* (7) *for ML2R and* (8) *for MLMC. Then*

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

  where the constant C_β reads

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

  and

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

We notice that for $\beta \geq 1$ the asymptotic behavior 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 $(1 + \frac{1}{2\alpha})$, whereas when $\beta < 1$, the asymptotic of the weights has an impact on the behavior of $N(\varepsilon)$ for ML2R. Still, in this case we observe that if $a_\infty = 1$ and $|\sum_{\ell=0}^{j-1} b_\ell| = 1$ 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 $(1 + \frac{1}{2\alpha})$ appears again to be the only difference in the coefficient C_β of $N(\varepsilon)$ for the two estimators.

Proof. ML2R: The estimator size N reads

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

We notice that $R(\varepsilon) \to +\infty$ as $\varepsilon \to 0$ and use Lemma 4.3 (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

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

as desired. □

5 Proofs

We will use the notations

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

where we set

$$\widetilde{Y}_j := Y^{(j)}_{\frac{\mathbf{h}}{n_j}} - Y^{(j)}_{\frac{\mathbf{h}}{n_{j-1}}} - \mathbb{E}\big[Y^{(j)}_{\frac{\mathbf{h}}{n_j}} - Y^{(j)}_{\frac{\mathbf{h}}{n_{j-1}}}\big], \quad j = 1, \dots, R(\varepsilon).$$

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

$$I^N_\pi(\varepsilon) - I_0 = \widetilde{I}^1_\varepsilon + \widetilde{I}^2_\varepsilon + \mu(\mathbf{h}, R(\varepsilon), M), \tag{22}$$

where the bias $\mu(\mathbf{h}, R(\varepsilon), M) \to 0$ as $\varepsilon \to 0$ (see Section 4.2 for a detailed description of the bias).

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

$$\mathbb{E}\big[|\widetilde{I}^2_\varepsilon|^p\big] \leq K(M, \beta, p)\varepsilon^p. \tag{23}$$

Proof. ML2R: We first give the proof of (23) for the ML2R estimator. As a first step we show that, for all $p \geq 2$,

$$\mathbb{E}\big[|\widetilde{Y}_j|^p\big] \leq C_{M,\beta,p} M^{-\frac{\beta p}{2}(j-1)}, \quad j = 1, \dots, R(\varepsilon), \quad \text{with } C_{M,\beta,p} = 2^p V_1^{(p)} (1 + M^{\frac{\beta}{2}})^p \mathbf{h}^{\frac{\beta p}{2}}. \tag{24}$$

By Minkowski's inequality,

$$\big(\mathbb{E}\big[|\widetilde{Y}_j|^p\big]\big)^{\frac{1}{p}} \leq \|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\|_p + \big|\mathbb{E}\big[Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\big]\big| \leq \|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\|_p + \|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\|_1 \leq 2\|Y_{\frac{\mathbf{h}}{n_j}} - Y_{\frac{\mathbf{h}}{n_{j-1}}}\|_p.$$

Applying again Minkowski's inequality, the L^p-strong approximation assumption (9) yields (24). As the random variables $(\widetilde{Y}^k_j)_{k\geq 1}$ are i.i.d. and the $(\widetilde{Y}_j)_{j=1,\dots,R(\varepsilon)}$ are centered and independent, Burkholder's inequality (see [10, Theorem 2.10, p. 23]) and (24) imply that there exists a positive universal real constant C_p such that

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

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 (20) and from inequality (21) 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

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

Owing to Lemma 4.5, up to reducing $\bar{\varepsilon}$, we have

$$\frac{1}{N(\varepsilon)} \leq \frac{2}{C_\beta} \varepsilon^2 \begin{cases} 1 & \text{if } \beta > 1, \\ R(\varepsilon)^{-1} & \text{if } \beta = 1, \\ M^{-\frac{1-\beta}{2} R(\varepsilon)} & \text{if } \beta < 1, \end{cases} \quad \text{for all } \varepsilon \in (0, \bar{\varepsilon}]. \tag{25}$$

Moreover,

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

Then

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

with

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

Hence (23) holds with

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

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

The Strong Law of Large Numbers follows as a consequence of Proposition 5.1.

Proof of Theorem 3.1. Owing to the decomposition (22), equation (10) amounts to proving

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

As $\lim_{\varepsilon\to 0} N_1(\varepsilon) = +\infty$ and the $(Y_{\mathbf{h}}^{(1),k})_{k\ge 1}$ 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 Lemma 5.1 it is straightforward that for all sequence of positive values $(\varepsilon_k)_{k\ge 1}$ such that $\varepsilon_k \to 0$ as $k \to +\infty$ and $\sum_{k\ge 1}\varepsilon_k^p < +\infty$,

$$\sum_{k\ge 1}\mathbb{E}[|\widetilde{I}^2_{\varepsilon_k}|^p] < +\infty.$$

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

5.2 Proof of the Central Limit Theorem

This subsection is devoted to the proof of Theorems 3.2 and 3.3. In order to satisfy a Lindeberg condition, we will need the assumption $(Z(h))_{h\in\mathcal{H}}$ is L^2-uniformly integrable. Owing to $(\mathrm{WE}_{\alpha,\bar{R}})$, $\bar{R} = 1$,

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

Since $2\alpha \ge \beta$, this deterministic sequence $(\mathbb{E}[Z(h)])_{h\in\mathcal{H}}$ 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 $(\widetilde{Z}(h))_{h\in\mathcal{H}} = (Z(h) - \mathbb{E}[Z(h)])_{h\in\mathcal{H}}$.

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

Lemma 5.2. *The following statements hold.*

(a) *If there exists a* $p > 2$ *such that* $\sup_{h\in\mathcal{H}}\|Z(h)\|_p < +\infty$, *the family* $(Z(h))_{h\in\mathcal{H}}$ *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* [3, Theorem 3.6]*):*
   (i) *The family* $(Z(h))_{h\in\mathcal{H}}$ *is* L^2*-uniformly integrable.*
   (ii) $\lim_{h\to 0}\|Z(h)\|_2 = \|D^{(M)}\|_2$.

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 3.2 and 3.3. Owing to the decomposition (22) (with $\mathbf{W}_j^{R(\varepsilon)} = 1$, $j = 1, \dots, R(\varepsilon)$ for MLMC estimator),

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

where $\widetilde{I}_\varepsilon^1$ and $\widetilde{I}_\varepsilon^2$ are independent. The bias term has already been treated in (17).

ML2R: Formulas (12) and (14) amount to proving, as $\varepsilon \to 0$,

$$\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1 \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\operatorname{Var}(Y_{\mathbf{h}})}{\mathbf{q}_1}\Big) \tag{26}$$

and

$$\frac{\widetilde{I}_\varepsilon^2}{\varepsilon} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_2^2) \tag{27}$$

with $\sigma_2 = \sigma$ for $\beta \in (0, 1]$. Indeed, for (26) let us write $\frac{\widetilde{I}_\varepsilon^1}{\varepsilon} = \frac{1}{\varepsilon\sqrt{N(\varepsilon)}}\sqrt{N(\varepsilon)}\widetilde{I}_\varepsilon^1$. Using Lemma 4.5, $N(\varepsilon)$ reads

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

In particular, since $R(\varepsilon) \to +\infty$ as $\varepsilon \to 0$, when $\beta \le 1$, $\frac{1}{\sqrt{N(\varepsilon)}} = o(\varepsilon)$ and the term $\frac{\widetilde{I}_\varepsilon^1}{\varepsilon} \to 0$ 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 behavior of the first term is driven by a regular Central Limit Theorem at rate $\sqrt{N(\varepsilon)}$, i.e.

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

which proves (26). We will use Lindeberg's Theorem for triangular arrays of martingale increments (see [10, Corollary 3.1, p. 58]) to establish (27). The random variables $\widetilde{Y}_j^k$ being centered and independent, the variance reads

$$\operatorname{Var}\Bigg(\sum_{j=2}^{R(\varepsilon)} \sum_{k=1}^{N_j(\varepsilon)} \frac{1}{\varepsilon} \frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)} \widetilde{Y}_j^k\Bigg) = \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \Bigg(\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\Bigg)^2 N_j(\varepsilon) \operatorname{Var}(\widetilde{Y}_j) = \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{N_j(\varepsilon)} \operatorname{Var}(\widetilde{Y}_j).$$

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

$$\Bigg|\frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{N_j(\varepsilon)} \operatorname{Var}(\widetilde{Y}_j) - \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{q_j(\varepsilon)N(\varepsilon)} \operatorname{Var}(\widetilde{Y}_j)\Bigg| \le \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{(q_j(\varepsilon)N(\varepsilon))^2} \operatorname{Var}(\widetilde{Y}_j).$$

The conclusion will follow from

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{N(\varepsilon)q_j(\varepsilon)} \operatorname{Var}(\widetilde{Y}_j) = \sigma_2^2 \tag{28}$$

and

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \sum_{j=2}^{R(\varepsilon)} \frac{(\mathbf{W}_j^{R(\varepsilon)})^2}{(q_j(\varepsilon)N(\varepsilon))^2} \operatorname{Var}(\widetilde{Y}_j) = 0. \tag{29}$$

Owing to the definition of Z_j given in (11), we get $\operatorname{Var}(\widetilde{Y}_j) = (\frac{\mathbf{h}}{n_j})^\beta \operatorname{Var}(Z_j)$ and, using the expression of $q_j(\varepsilon)$ given in (20), we obtain

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

Case $\beta > 1$: Owing to the expression of $N(\varepsilon)$ given in Lemma 4.5 when $\beta > 1$,

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

and owing to the limit in Lemma 4.3 (d) with $\gamma = \frac{1-\beta}{2} < 0$,

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

Hence the convergence of the variance (28) holds for Theorem 3.2.

Case $\beta \le 1$: Owing to the expression of $N(\varepsilon)$ given in Lemma 4.5 when $\beta \le 1$, we get, as $\varepsilon \to 0$,

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

We notice that

$$\lim_{j\to+\infty} \operatorname{Var}(Z_j) = \tilde{v}_\infty(M,\beta)$$

if $2\alpha > \beta$ and

$$\lim_{j\to+\infty} \operatorname{Var}(Z_j) = \tilde{v}_\infty(M,\beta) - c_1^2(1-M^{\frac{\beta}{2}})^2$$

if $2\alpha = \beta$. Hence, owing to the limit in Lemma 4.3 (d) with $\gamma = \frac{1-\beta}{2} \ge 0$, we obtain (28) with $\sigma_2 = \sigma$ given in (13) in Theorem 3.3.

For (29), it follows from the expression of $q_j(\varepsilon)$ in (20) 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 (11) and to inequality (21), we get

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

We conclude by showing that

$$\frac{M^{R(\varepsilon)}}{(\varepsilon N(\varepsilon))^2} \to 0 \quad \text{as } \varepsilon \to 0. \tag{30}$$

Owing to the expression of $R(\varepsilon)$ given in (7), we notice that

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

Moreover, using Lemma 4.5, up to another reduction of $\bar{\varepsilon}$, we have $\frac{1}{(\varepsilon N(\varepsilon))^2} \le (\frac{2}{C_\beta})^2 \varepsilon^2$ for all $\beta > 0$. This in turn yields

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

Then (29) 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 $\eta > 0$,

$$\sum_{j=2}^{R(\varepsilon)} \sum_{k=1}^{N_j(\varepsilon)} \mathbb{E}\Big[\Big(\frac{1}{\varepsilon} \frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)} \widetilde{Y}_j^k\Big)^2 \mathbf{1}_{\big\{\big|\frac{1}{\varepsilon} \frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)} \widetilde{Y}_j^k\big| > \eta\big\}}\Big] \to 0 \quad \text{as } \varepsilon \to 0.$$

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

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

We set $\widetilde{Z}_j = Z_j - \mathbb{E}[Z_j]$. Replacing q_j by its values given in (20), using inequality (19) from Lemma 4.3 (b) and the elementary inequality $\frac{1}{N_j(\varepsilon)} \le \frac{1}{q_j(\varepsilon)N(\varepsilon)}$ yields

$$\begin{aligned}
&\sum_{j=2}^{R(\varepsilon)} \sum_{k=1}^{N_j(\varepsilon)} \mathbb{E}\Big[\Big(\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\Big)^2 \mathbf{1}_{\big\{\big|\frac{1}{\varepsilon}\frac{\mathbf{W}_j^{R(\varepsilon)}}{N_j(\varepsilon)}\widetilde{Y}_j^k\big|>\eta\big\}}\Big] \\
&\quad \le \frac{1}{\theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta}\mu^*(\varepsilon)} \frac{1}{\varepsilon^2 N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} |\mathbf{W}_j^{R(\varepsilon)}| M^{\frac{\beta+1}{2}(j-1)} \mathbb{E}\Big[(\widetilde{Y}_j)^2 \mathbf{1}_{\Big\{|\widetilde{Y}_j|>\eta \frac{\theta \mathbf{h}^{\frac{\beta}{2}} \underline{C}_{M,\beta}\mu^*(\varepsilon)}{M^{\frac{\beta+1}{2}(j-1)}}\varepsilon N(\varepsilon)\Big\}}\Big] \\
&\quad \le \frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta \underline{C}_{M,\beta}\underline{\mu}^*} a_\infty \widetilde{B}_\infty \frac{1}{\varepsilon^2 N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)} \mathbb{E}\Big[(\widetilde{Z}_j)^2 \mathbf{1}_{\{|\widetilde{Z}_j|>\eta\theta \underline{C}_{M,\beta}\underline{\mu}^* \varepsilon N(\varepsilon) M^{-\frac{j-1}{2}}\}}\Big] \\
&\quad \le \frac{\mathbf{h}^{\frac{\beta}{2}}}{\theta \underline{C}_{M,\beta}\underline{\mu}^*} a_\infty \widetilde{B}_\infty \sup_{2\le j\le R(\varepsilon)} \mathbb{E}\Big[(Z_j)^2 \mathbf{1}_{\{|\widetilde{Z}_j|>\Theta\varepsilon N(\varepsilon) M^{-\frac{R(\varepsilon)}{2}}\}}\Big] \frac{1}{\varepsilon^2 N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)}
\end{aligned}$$

where we set $\Theta = \eta\theta \underline{C}_{M,\beta}\underline{\mu}^* \sqrt{M}$. Now, it follows from Lemma 4.5 that

$$\frac{1}{\varepsilon^2 N(\varepsilon)} \sum_{j=2}^{R(\varepsilon)} M^{\frac{1-\beta}{2}(j-1)} = \frac{1}{\varepsilon^2 N(\varepsilon)} \Big(\frac{M^{\frac{1-\beta}{2}R(\varepsilon)} - M^{\frac{1-\beta}{2}}}{M^{\frac{1-\beta}{2}} - 1} \mathbf{1}_{\{\beta\ne 1\}} + (R(\varepsilon)+1)\mathbf{1}_{\{\beta=1\}} \Big) \to K,$$

as $\varepsilon \to 0$, where K is a real positive constant. Owing to (30), $\lim_{\varepsilon\to 0} \varepsilon N(\varepsilon) M^{-\frac{R(\varepsilon)}{2}} = +\infty$. Hence, since we assumed that the family $(Z_j)_{j\ge 1}$ is L^2-uniformly integrable, we obtain that

$$\lim_{\varepsilon\to 0} \sup_{2\le j\le R(\varepsilon)} \mathbb{E}\Big[(Z_j)^2 \mathbf{1}_{\{|Z_j|>\Theta\varepsilon N(\varepsilon) M^{-\frac{R(\varepsilon)-1}{2}}\}}\Big] = 0$$

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 4.5 and the expression of $R(\varepsilon)$ given in (8), we obtain

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

We replace $\mathbf{W}_j^R = 1$, $j = 1, \ldots, R$, and $a_\infty \widetilde{B}_\infty = 1$. The only significant difference comes when $\beta < 1$, while proving (30). In this case, owing to Lemma 4.5 as we did in (25) and using the expression of $R(\varepsilon)$ given in (8), up to reducing $\bar{\varepsilon}$, we can write

$$\frac{M^{R(\varepsilon)}}{(\varepsilon N(\varepsilon))^2} \le \Big(\frac{2}{C_\beta}\Big)^2 \varepsilon^2 M^{-(1-\beta)R(\varepsilon)} M^{R(\varepsilon)} \le \Big(\frac{2}{C_\beta}\Big)^2 M^{\beta(C_R^{(1)}+1)} \varepsilon^{2-\frac{\beta}{\alpha}},$$

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

6 Applications

6.1 Diffusions

In this subsection we retrieve a recent result by Kebaier and Ben Alaya (see [2]) 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

$$X_t = X_0 + \int_0^t b(s, X_s)\,ds + \int_0^t \sigma(s, X_s)\,dW_s, \quad t \in [0, T],$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to M(d, q, \mathbb{R})$ 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 $(\Omega, \mathcal{A}, \mathbb{P})$.

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

$$\bar{X}_t^n = X_0 + \int_0^t b(\underline{s}, \bar{X}_{\underline{s}}^n)\, ds + \int_0^t \sigma(\underline{s}, \bar{X}_{\underline{s}}^n)\, dW_s \tag{31}$$

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

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

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 (31) 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:

$$\forall\, p \geqslant 2,\ \exists\, c_{b,\sigma,p,T} > 0, \quad \Big\| \sup_{t\in[0,T]} |X_t - \bar{X}_t^n| \Big\|_p \leq c_{b,\sigma,p,T} \sqrt{\frac{T}{n}}(1 + \|X_0\|_p). \tag{32}$$

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 6.1. *The following statements hold.*

(a) Regular setting (Talay–Tubaro [13]): *If b and σ are infinitely differentiable with bounded partial derivatives and if $f : \mathbb{R}^d \to \mathbb{R}$ 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 $R \in N^*$,*

$$\mathbb{E}[f(\bar{X}_T^n)] - \mathbb{E}[f(X_T)] = \sum_{k=1}^{R} c_k\Big(\frac{1}{n}\Big)^k + O\Big(\Big(\frac{1}{n}\Big)^{R+1}\Big), \tag{33}$$

where the coefficients c_k depend on b, σ, f, T but not on n.

(b) (Hypo-)Elliptic setting (Bally–Talay [1]): *If b and σ are infinitely differentiable with bounded partial derivatives and if σ is uniformly elliptic in the sense that*

$$\sigma\sigma^*(x) \geq \varepsilon_0 I_q \quad \textit{for all } x \in \mathbb{R}^d \textit{ and all } t \in [0, T], \quad \varepsilon_0 > 0,$$

or more generally if (b, σ) satisfies the strong Hörmander hypo-ellipticity assumption, then (33) *holds true for every bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$.*

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^α, $\alpha \in (0, 1]$, have sometimes been established (see [12] for a brief review in connection with multilevel methods). However, as emphasized by the numerical experiments carried out in [12], such weak error expansion can highly be suspected to hold at any order under reasonable smoothness assumptions.

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

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

We assume the weak error expansion $(\mathrm{WE}_{\alpha,\bar{R}})$. We prove now that both estimators ML2R (2) and MLMC (1) satisfy a Strong Law of Large Numbers and a Central Limit Theorem when ε tends to 0.

Theorem 6.2. *Let $X_0 \in L^2$ and assume that $F : \mathcal{C}_b([0,T], \mathbb{R}^d) \to \mathbb{R}$ is a Lipschitz continuous functional. Then the assumption* (SE$_\beta$) *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 3.1.

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

$$\exists\, v_\infty > 0, \quad \lim_{h\to 0} \|Z(h)\|_2^2 = (M-1)v_\infty. \tag{34}$$

As a consequence, both ML2R and MLMC estimators satisfy Theorem 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 $p \geqslant 2$,

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

then $(Y_h)_{h\in\mathcal{H}}$ satisfies (SE$_\beta$) 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 criterion (a) of Lemma 5.2, we prove that $(Z(h))_{h\in\mathcal{H}}$ is L^2-uniformly integrable.

At this stage it remains to prove (34). The key is [2, Theorem 3], where it is proved that

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

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

$$U_t^{(M)} = \sqrt{\frac{M-1}{2}} \sum_{i,j=1}^{q} V_t \int_0^t (V_s)^{-1} \nabla\varphi_{.j}(X_s)\varphi_{.i}(X_s)\, dB_s^{i,j}, \quad t \in [0,T]. \tag{35}$$

We recall the notations of Jacod and Protter [11]

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

with $\varphi_{.j}$ representing the jth column of the matrix $\varphi = [\varphi_{ij}]_{i=1,\dots,d,\, j=1,\dots,q}$, 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 $\mathbb{R}^{d\times d}$ valued process solution of the linear equation

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

Here $(B^{ij})_{1\leq i,j\leq q}$ is a standard q^2-dimensional Brownian motion independent of W. This process is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbb{P}})$ of the original space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ on which lives W.

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

$$Z(h) = \sqrt{nM}\big(F(\bar{X}^{nM}) - F(\bar{X}^n)\big) = -\int_0^1 DF(u\bar{X}^n + (1-u)\bar{X}^{nM})\, du \cdot U_n^{(M)},$$

where $U_n^{(M)} := \sqrt{nM}(\bar{X}^n - \bar{X}^{nM})$. The function $(x_1, x_2, x_3) \mapsto \int_0^1 DF(ux_1 + (1-u)x_2)\, dux_3$ is continuous, and it suffices to prove that $(\bar{X}^n, \bar{X}^{nM}, U_n^{(M)}) \xrightarrow{\mathcal{L}} (X, X, U^{(M)})$, as n goes to infinity, to conclude that

$$Z(h) \xrightarrow{\mathcal{L}} -DF(X)U^{(M)} \quad \text{as } h \to 0. \tag{36}$$

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

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

Since $(U_n^{(M)})_{n\geqslant 1}$ converges stably with limit $U^{(M)}$, we have that

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

On the other hand, owing to (32), we prove that

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

By (36) and Lemma 5.2 (b) we have

$$\lim_{h\to 0}\|Z(h)\|_2^2 = \|DF(X)U^{(M)}\|_2^2 = (M-1)v_\infty$$

with $v_\infty = \|DF(X)\frac{U^{(M)}}{M-1}\|_2^2$ which does not depend on M owing to the definition of U^M given in (35). □

6.2 Nested Monte Carlo

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

$$\mathbb{E}[f(\mathbb{E}[X\,|\,Y])],$$

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

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

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

$$Y_0 := f(\mathbb{E}[X\,|\,Y]), \quad Y_h = Y_{\frac{1}{K}} := f\bigg(\frac{1}{K}\sum_{k=1}^{K} F(\xi_k, Y)\bigg),$$

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 $(n_j = M^{j-1})$

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

where $(Y^{(j),i})_{i\geq 1}$ is a sequence of independent copies of $Y^{(j)} \sim Y$, $j = 1, \dots, R$, $Y^{(j)}$ independent of $Y^{(\ell)}$ for $j \neq \ell$, and $(\xi_k^{(j),i})_{k,i\geq 1,\, j=1,\dots,R}$ is a sequence of i.i.d. variables $\xi_k^{(j),i} \sim \xi$. We saw in [12] that, when f is $2R$ times differentiable with $f^{(k)}$ bounded, the *nested* Monte Carlo estimator satisfies (SE_β) with $\beta = 1$ and $(\mathrm{WE}_{\alpha,\bar{R}})$ 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 3.1 and of the Central Limit Theorem 3.3. Then we define for convenience

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

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

Proposition 6.3. *Still assuming that f is Lipschitz continuous. If $X \in L^p(\mathbb{P})$ for $p \geqslant 2$, then there exists $V_1^{(p)}$ such that, for all $h = \frac{1}{K}$ and $h' = \frac{1}{K'}$, $K, K' \in K_0\mathbb{N}^*$,*

$$\|Y_{h'} - Y_h\|_p^p \leq V_1^{(p)}|h' - h|^{\frac{p}{2}}. \tag{37}$$

As a consequence, assumption (SE_β) and the L^p-strong error assumption (9) are satisfied with $\beta = 1$. Then both ML2R and MLMC estimators satisfy a Strong Law of Large Numbers, see Theorem 3.1.

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

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

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

$$\mathbb{E}\left[\left|\frac{S_{K'}}{K'} - \frac{S_K}{K}\right|^p\right] = \mathbb{E}\left[\left|\left(\frac{1}{K'} - \frac{1}{K}\right)S_K + \frac{1}{K'}(S_{K'} - S_K)\right|^p\right]$$
$$\leq 2^{p-1}\left[\left|\frac{1}{K'} - \frac{1}{K}\right|^p \mathbb{E}[|S_K|^p] + \left(\frac{1}{K'}\right)^p \mathbb{E}[|S_{K'} - S_K|^p]\right].$$

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

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

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

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

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

$$\left|\frac{1}{K'} - \frac{1}{K}\right|^p K^{\frac{p}{2}} + \left|\frac{1}{K'}\right|^p |K' - K|^{\frac{p}{2}} = \left|\frac{1}{K'} - \frac{1}{K}\right|^{\frac{p}{2}}\left|\frac{K}{K'} - 1\right|^{\frac{p}{2}} + \left|\frac{K}{K'}\right|^{\frac{p}{2}}\left|\frac{1}{K} - \frac{1}{K'}\right|^{\frac{p}{2}} \leq 2\left|\frac{1}{K} - \frac{1}{K'}\right|^{\frac{p}{2}}.$$

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

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

Lemma 6.4. *Assume that $f : \mathbb{R} \to \mathbb{R}$ 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 $h \to 0$,*

$$Z(h) = \sqrt{\frac{M}{h}}(Y_{\frac{h}{M}} - Y_h) \xrightarrow{\mathcal{L}} \sqrt{M-1} f'(\phi_0(Y))\sigma_F(Y)\zeta. \tag{38}$$

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

$$z_h^{(M)}(y) = \sqrt{\frac{M}{h}}(f(\phi_{\frac{h}{M}}(y)) - f(\phi_h(y))) \quad \text{for all } y \in \mathbb{R}^{q_Y}.$$

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

$$z_h^{(M)}(y) = -\left(\int_0^1 f'(v\phi_h(y) + (1-v)\phi_{\frac{h}{M}}(y))\, dv\right) u_h^{(M)}(y) \tag{39}$$

with

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

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 $(x_1, x_2) \mapsto \int_0^1 f'(vx_1 + (1-v)x_2)\, dv$ (since f' is continuous) we get

$$\lim_{h\to 0} \int_0^1 f'(v\phi_h(y) + (1-v)\phi_{\frac{h}{M}}(y))\, dv = f'(\phi_0(y)) \quad \text{a.s.} \tag{40}$$

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

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

Owing to the Central Limit Theorem and the independence of both terms on 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 $(\frac{M-1}{\sqrt{M}})^2 + (\sqrt{\frac{M-1}{M}})^2 = M-1$, we obtain

$$u_h^{(M)}(y) \xrightarrow{\mathcal{L}} \sqrt{M-1}\sigma_F(y)\zeta \quad \text{with } \zeta \sim \mathcal{N}(0,1). \tag{41}$$

By Slutsky's Theorem, we derive from (39), (40) and (41) that for every $y \in \mathbb{R}^{q_Y}$,

$$z_h^{(M)}(y) \xrightarrow{\mathcal{L}} \sqrt{M-1}f'(\phi_0(y))\sigma_F(y)\zeta \quad \text{as } h \to 0. \tag{42}$$

Recall that $Z(h) = z_h^{(M)}(Y)$. We prove (38) combining Fubini's Theorem with the Lebesgue Dominated Convergence Theorem and (42). More precisely, for all $G \in \mathcal{C}_b$ we have

$$\begin{aligned}\lim_{h\to 0}\mathbb{E}[[]2]G(Z(h)) &= \lim_{h\to 0}\mathbb{E}[G(z_h^{(M)}(Y))] = \mathbb{E}[\lim_{h\to 0} G(z_h^{(M)}(Y))]\\ &= \mathbb{E}[G(\sqrt{M-1}f'(\phi_0(y))\sigma_F(y)\zeta)],\end{aligned}$$

as desired. □

We are now in a position to prove that the *nested* Monte Carlo satisfies the assumptions of the Central Limit Theorem 3.3.

Theorem 6.5. *Assume that $f : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function and differentiable with f' continuous. Then $(Z(h))_{h\in\mathcal{H}}$ is L^2-uniformly integrable and*

$$\lim_{h\to 0}\|Z(h)\|_2^2 = (M-1)\|f'(\phi_0(Y))\sigma_F(Y)\|_2^2. \tag{43}$$

As a consequence, the ML2R and MLMC estimators (2) *and* (1) *satisfy a Central Limit Theorem in the sense of Theorem 3.3 (case $\beta = 1$).*

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

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

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

We saw in the proof of Proposition 6.4 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 5.2 (b), the uniform integrability will follow from $\lim_{h\to 0}\|u_h^{(M)}(Y)\|_2 = \|\sqrt{M-1}\sigma_F(Y)\zeta\|_2$. In fact, this convergence holds as an equality. Indeed,

$$\|u_h^{(M)}(Y)\|_2^2 = MK\mathbb{E}[(S_K - S_{MK})^2] = MK\mathbb{E}\Big[\Big(\frac{M-1}{MK}S_K - \frac{1}{MK}(S_{MK} - S_K)\Big)^2\Big].$$

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

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

We prove now (43) using again Lemma 5.2 (b) with the convergence in law of $(Z(h))_{h\in\mathcal{H}}$ established in Lemma 6.4. □

We notice that, if assumption (37) in Proposition 6.3 holds with $p > 2$, the condition of L^2-uniform integrability is much easier to show since it is a direct consequence of Lemma 5.2 (a).

6.3 Smooth nested Monte Carlo

When the function f is smooth, namely $\mathcal{C}^{1+\rho}(\mathbb{R},\mathbb{R})$, $\rho \in (0,1]$ (f' is ρ-Hölder), a variant of the former multilevel nested estimator has been used in [4] (see also [8]) to improve the strong rate of convergence in order to attain the asymptotically unbiased setting $\beta > 1$ in 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 $h = \frac{1}{K}$, $K \in K_0\mathbb{N}^*$) in the ML2R and MLMC estimators by

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

It is clear that

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

Computations similar to those carried out in Proposition 6.3 yield that, if $X = F(\xi, Y) \in L^{p(1+\rho)}(\mathbb{P})$ for some $p \geqslant 2$, then

$$\|Y_{h,\frac{h}{M}}\|_p^p \leqslant V_M^{(\rho,p)}\Big|h - \frac{h}{M}\Big|^{\frac{p}{2}(1+\rho)} = V_M^{(\rho,p)}\Big|1 - \frac{1}{M}\Big|^{\frac{p}{2}(1+\rho)}|h|^{\frac{p}{2}(1+\rho)}. \tag{44}$$

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 3.1.

CLT: When (44) 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 3.2) is the L^2-uniform integrability of $(h^{-\frac{\beta}{2}}Y_{h,\frac{h}{M}})_{h\in\mathcal{H}}$, since no sharp rate is needed when $\beta > 1$. Moreover, if (44) holds for a $p \in (2, +\infty)$, i.e. if $X = F(\xi, Y) \in L^{p(1+\rho)}(\mathbb{P})$ with $p > 2$, then

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

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 $\mathbb{E}[Y_{\frac{h}{M}}] \leqslant \mathbb{E}[Y_h]$ which in turn implies by an easy induction that $\mathbb{E}[Y_0] \leqslant \mathbb{E}[Y_h]$ 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 [9].

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 diffusions, where we extend a result of Ben Alaya and Kebaier in [2], and the nested Monte Carlo, first mentioned in the Multilevel framework by Lemaire and Pagès in [12]. The Strong Law of Large Numbers is essentially a consequence of the strong error assumption (SE_β) (or of its reinforced version (9)), and of the estimator levels' independence. The understanding of the behavior of the weights 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 follow 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 1 and 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.

A Asymptotic of the weights

We focus our attention on the behavior 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 $j \ge R + 1$, $R \in \mathbb{N}^*$. 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 4.3 is then proved. As a consequence,

$$|\mathbf{W}_j^R| \le a_\infty \widetilde{B}_\infty \quad \text{for all } R \in \mathbb{N}^* \text{ and all } j \in \{1, \dots, R\},$$

which proves claim (b) in Lemma 4.3. For the proof of claims (c) and (d), we will need the following:

Lemma A.1. *Let $\varphi : \mathbb{N}^* \to \mathbb{N}^*$ such that $\varphi(R) \in \{1, \dots, R-1\}$ for all $R \ge 1$, $\varphi(R) \to +\infty$ and $R - \varphi(R) \to +\infty$ as $R \to +\infty$. Then*

$$\lim_{R\to+\infty} \sup_{1\le j\le\varphi(R)} |\mathbf{W}_j^R - 1| = 0.$$

In particular, for all $j \in \mathbb{N}^*$ we have $\mathbf{W}_j^R \to 1$ 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 $j \in \mathbb{N}^*$ as $R \to +\infty$.

Proof. We write

$$|\mathbf{W}_j^R - a_\infty B_\infty| = \left| \sum_{\ell=0}^{R-j} a_{R-\ell} b_\ell - a_\infty \sum_{\ell=0}^{R-j} b_\ell - a_\infty \sum_{\ell\ge R-j+1} b_\ell \right| \le \sum_{\ell=0}^{R-j} (a_\infty - a_{R-\ell}) |b_\ell| + a_\infty \sum_{\ell\ge R-j+1} |b_\ell|.$$

First note that

$$\lim_{R\to+\infty} \sup_{j\in\{1,\ldots,\varphi(R)\}} \sum_{\ell\geq R-j+1} |b_\ell| \leq \lim_{R\to+\infty} \sum_{\ell\geq R-\varphi(R)+1} |b_\ell| = 0$$

as $R-\varphi(R)\to+\infty$ and $\sum_{\ell\geq 0}|b_\ell|<+\infty$. On the other hand, for every $j\in\{1,\ldots,\varphi(R)\}$,

$$\begin{aligned}
\sum_{\ell=0}^{R-j}(a_\infty - a_{R-\ell})|b_\ell| &= \sum_{\ell=j}^{R}(a_\infty - a_\ell)|b_{R-\ell}| \\
&= \sum_{\ell=j}^{\varphi(R)}(a_\infty - a_\ell)|b_{R-\ell}| + \sum_{\ell=\varphi(R)+1}^{R}(a_\infty - a_\ell)|b_{R-\ell}| \\
&\leq a_\infty \sum_{\ell=R-\varphi(R)}^{R-j}|b_\ell| + (a_\infty - a_{\varphi(R)+1})\sum_{\ell=0}^{R-\varphi(R)-1}|b_\ell| \\
&\leq a_\infty \sum_{\ell=R-\varphi(R)}^{+\infty}|b_\ell| + (a_\infty - a_{\varphi(R)+1})\sum_{\ell=0}^{R-\varphi(R)-1}|b_\ell|.
\end{aligned}$$

Consequently, $\sup_{j\in\{1,\ldots,R\}}\sum_{\ell=0}^{R-j}(a_\infty - a_{R-\ell})|b_\ell|\to 0$ as $R\to+\infty$, since $\varphi(R)$ and $R-\varphi(R)\to+\infty$. Finally,

$$\lim_{R\to+\infty}\sup_{1\leq j\leq\varphi(R)}|\mathbf{W}_j^R - a_\infty B_\infty| = 0.$$

Moreover, by definition we have $\mathbf{W}_1^R = 1$ for all R, which implies that $B_\infty = \frac{1}{a_\infty}$ and completes the proof. Finally, as $a_j\to a_\infty$,

$$\mathbf{W}_{R-j+1}^R = \sum_{\ell=0}^{j-1} a_{R-\ell} b_l \xrightarrow{R\to+\infty} a_\infty \sum_{\ell=0}^{j-1} b_\ell,$$

as desired. □

Proof of Lemma 4.3 (c) *and* (d). (c) Let us consider the non-negative measure on $\mathbb{N}^*$ defined by

$$m_\beta(j) = M^{\gamma(j-1)}, \quad \gamma<0.$$

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.1, $\mathbf{W}_j^R\to 1$ as $R\to+\infty$ for every $j\in\mathbb{N}^*$ and $|\mathbf{W}_j^R|\leq a_\infty\widetilde{B}_\infty$, we derive from Lebesgue's Dominated Convergence Theorem that

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

(d) If $\gamma<0$, we consider the non-negative finite measure on $\mathbb{N}^*$ defined by $m'_\beta(j) = M^{\gamma(j-1)}v_j$ since $(v_j)_{j\geq 1}$ is a bounded sequence of positive real numbers. As in the previous case (c) we have

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

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

$$\begin{aligned}
\left|\frac{1}{R}\sum_{j=2}^{R}|\mathbf{W}_j^R|v_j - \frac{1}{R}\sum_{j=2}^{R}v_j\right| &\leq \left[\frac{1}{R}\sum_{j=2}^{\varphi(R)}\big||\mathbf{W}_j^R|-1\big| + \frac{1}{R}\sum_{j=\varphi(R)+1}^{R}(|\mathbf{W}_j^R|+1)\right]\sup_{j\geq 2}v_j \\
&\leq \left[\sup_{2\leq j\leq\varphi(R)}|\mathbf{W}_j^R-1|\frac{\varphi(R)}{R} + (a_\infty\widetilde{B}_\infty+1)\Big(1-\frac{\varphi(R)}{R}\Big)\right]\sup_{j\geq 2}v_j.
\end{aligned}$$

Owing to Lemma A.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

$$\sum_{j=2}^{R} |\mathbf{W}_j^R| M^{\gamma(j-1)} \geq |\mathbf{W}_R^R| M^{\gamma R} = |a_R| M^{\gamma R} \to +\infty. \tag{45}$$

Let $\eta > 0$. Since $\lim_{j\to+\infty} v_j = 1$, there exists $N_\eta \in \mathbb{N}^*$ such that, for each $j > N_\eta$, $|v_j - 1| < \frac{\eta}{2}$. Owing to Lemma A.1 there exists R_η such that, for each $R \geq R_\eta$, $\sup_{2\leq j\leq N_\eta} |\mathbf{W}_j^R| < 1 + \eta$. Then

$$\begin{aligned}
\left| \frac{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)} v_j}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} - 1 \right| &\leq \frac{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)} |v_j - 1|}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} \\
&\leq \frac{\sum_{j=2}^{N_\eta} |\mathbf{W}_j^R| M^{\gamma(j-1)} |v_j - 1|}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} + \frac{\eta}{2} \frac{\sum_{j=N_\eta+1}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} \\
&\leq \frac{\max_{2\leq j\leq N_\eta} M^{\gamma(j-1)} |v_j - 1| N_\eta \sup_{2\leq j\leq N_\eta} |\mathbf{W}_j^R|}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} + \frac{\eta}{2} \\
&\leq \frac{f(N_\eta)(1+\eta)}{\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)}} + \frac{\eta}{2}
\end{aligned}$$

where $f(N) = \max_{2\leq j\leq N} M^{\gamma(j-1)} |v_j - 1| N$ does not depend on R. Thanks to (45), there exists $R'_\eta > 0$ such that, for each $R \geq \max(R_\eta, R'_\eta)$, $\sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)} > \frac{2f(N_\eta)(1+\eta)}{\eta}$, which proves that

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

This leads to analyze

$$\frac{1}{M^{\gamma R}} \sum_{j=2}^R |\mathbf{W}_j^R| M^{\gamma(j-1)} = \sum_{j=2}^R |\mathbf{W}_j^R| M^{-\gamma(R-j+1)} = \sum_{j=1}^{R-1} |\mathbf{W}_{R-j+1}^R| M^{-\gamma j}.$$

Using that $|\mathbf{W}_j^R| \leq a_\infty \widetilde{B}_\infty$ for $j \in \{1, \dots, R\}$ and Lemma A.1, one derives from Lebesgue's Dominated Convergence Theorem that

$$\sum_{j=1}^{R-1} |\mathbf{W}_{R-j+1}^R| M^{-\gamma j} \xrightarrow{R\to+\infty} a_\infty \sum_{j\geq 1} \left| \sum_{\ell=0}^{j-1} b_\ell \right| M^{-\gamma j} < +\infty$$

since $|\sum_{\ell=0}^{j-1} b_\ell| \leq \sum_{\ell=0}^{j-1} |b_\ell| \leq \widetilde{B}_\infty$. □

References

[1] V. Bally and D. Talay, The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function, *Probab. Theory Related Fields* **104** (1996), no. 1, 43–60.

[2] M. Ben Alaya and A. Kebaier, Central limit theorem for the multilevel Monte Carlo Euler method, *Ann. Appl. Probab.* **25** (2015), no. 1, 211–234.

[3] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley Ser. Probab. Stat., John Wiley & Sons, New York, 1999.

[4] K. Bujok, B. Hambly and C. Reisinger, Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives, *Methodol. Comput. Appl. Probab.* **17** (2015), no. 3, 579–604.

[5] D. Duffie and P. Glynn, Efficient Monte Carlo simulation of security prices, *Ann. Appl. Probab.* **5** (1995), no. 4, 897–905.

[6] A. A. Gerbi, B. Jourdain and E. Clément, Ninomiya–Victoir scheme: Strong convergence, antithetic version and application to multilevel estimators, *Monte Carlo Methods Appl.* **22** (2016), no. 3, 197–228.

[7] M. Giles, Multilevel Monte Carlo path simulation, *Oper. Res.* **56** (2008), no. 3, 607–617.

[8] M. B. Giles, Multilevel Monte Carlo methods, *Acta Numer.* **24** (2015), 259–328.

[9] D. Giorgi, Ph.D. thesis, in progress.

[10] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Probab. Math. Statist., Academic Press, New York, 1980.

[11] J. Jacod and P. Protter, Asymptotic error distributions for the euler method for stochastic differential equations, *Ann. Probab.* **26** (1998), no. 1, 267–307.
[12] V. Lemaire and G. Pagès, Multilevel richardson-romberg extrapolation, preprint (2014), https://arxiv.org/abs/1401.1177v1; to appear in *Bernoulli*.
[13] D. Talay and L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, *Stoch. Anal. Appl.* **8** (1990), no. 4, 483–509.