Weak error for nested Multilevel Monte Carlo

Daphné Giorgi, Vincent Lemaire, Gilles Pagès

Weak error for nested Multilevel Monte Carlo

Daphné Giorgi[*], Vincent Lemaire[†], Gilles Pagès[‡]

June 17, 2018

Abstract

This article discusses MLMC estimators with and without weights, applied to nested expectations of the form $\mathbf{E}\left[f(\mathbf{E}\left[F(Y,Z)|Y\right])\right]$. More precisely, we are interested on the assumptions needed to comply with the MLMC framework, depending on whether the payoff function f is smooth or not. A new result to our knowledge is given when f is not smooth in the development of the weak error at an order higher than 1, which is needed for a successful use of MLMC estimators with weights.

Keywords: Multilevel Monte Carlo; Weighted Multilevel Monte Carlo; Nested Monte Carlo; Weak error expansion.
MSC 2010: primary 65C05; secondary 65C30.

1 Introduction

Multilevel estimators are commonly used when the underlying random variable of interest – here $f\left(\mathbf{E}\left[F(Y,Z)|Y\right]\right)$, with Y and Z independent as far as nested simulation is concerned – cannot be simulated exactly at a reasonable computational cost. However, such approximations – here $f\left(\frac{1}{N}\sum_{k=1}^{N} F(Y,Z_k)\right)$ – induce some bias. Nested simulation is one of the two most popular setting where Multilevel method are implemented, the other being the numerical schemes associated to stochastic dynamics.

The optimal calibration and the resulting performances of Multilevel Monte Carlo estimators depend on the weak and strong error rate of convergence of these simulable proxies toward $f\left(\mathbf{E}\left[F(Y,Z)|Y\right]\right)$. By weak error, we mean here an expansion of the bias as a function of a given parameter h representative of the (inverse) complexity.

The existence of weak error expansions at order one leads to the (regular and original) Multilevel Monte Carlo (MLMC) method introduced by M. Giles in [Gil08], whereas higher order expansions led naturally to develop a weighted multilevel framework, called Richardson-Romberg Multilevel method (ML2R) introduced in [LP17]. However, the existence of such an

[*]Sorbonne Université, Sorbonne Paris Cité, CNRS, Laboratoire de Probabilités Statistique et Modélisation, LPSM, F-75005 Paris, France, E-mail: daphne.giorgi@sorbonne-universite.fr

[†]Sorbonne Université, Sorbonne Paris Cité, CNRS, Laboratoire de Probabilités Statistique et Modélisation, LPSM, F-75005 Paris, France, E-mail: vincent.lemaire@sorbonne-universite.fr

[‡]Sorbonne Université, Sorbonne Paris Cité, CNRS, Laboratoire de Probabilités Statistique et Modélisation, LPSM, F-75005 Paris, France, E-mail: gilles.pages@sorbonne-universite.fr

expansion not only depends upon random variable of interest and its approximations but also on the regularity of the "payoff" function f, as it has been widely popularized by the analysis of time discretization schemes of Brownian diffusion processes (see [TT90] and [BT96a]).

The seminal result concerning the first order weak error expansion for nested Monte Carlo simulation when f is not regular – namely a quantile – is due to Gordy and Juneja in [GJ10].

For such indicator function the strong rate of convergence remains slow and reduces the efficiency of regular multilevel estimators since their performances are ruled by this strong convergence rate. In particular they no longer behave as unbiased or almost unbiased estimators as it is the case for faster strong convergence regimes.

By contrast weighted multilevel estimators are still almost unbiased in some sense but this performance strongly relies on higher order expansions of the weak error. So the main objective and result of this paper is to establish (see Proposition 5.1) such a higher order expansion for non-smooth payoff function f in a nested simulation framework.

Let us briefly recall the multilevel paradigm (see [Pag18]). Let $Y_0 \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ be a random variable and Y_h, $h \in \mathcal{H} = \left\{\frac{\mathbf{h}}{n},\, n \geqslant 1\right\}$ be a family of approximations of Y_0 such that $\lim_{h\to 0} \|Y_h - Y_0\|_2 = 0$ with a simulation cost of the form

$$\mathrm{Cost}(Y_h) = \kappa h^{-1}$$

so that the parameter h is inverse linear in the complexity. Its role in the weak expansion error will lead us to call it *bias parameter*.

The central idea behind the regular MLMC estimator is to consider a R-tuple of parameters $h_j = h/M^{j-1}$, $j = 1, \dots, R$ ($h \in \mathcal{H}$) and to write the telescopic sum

$$\mathbf{E}\left[Y_{h_R}\right] = \mathbf{E}\left[Y_{h_1}\right] + \sum_{j=2}^{R} \mathbf{E}\left[Y_{h_j} - Y_{h_{j-1}}\right]$$

which suggests to introduce the estimator (see [Gil08])

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

where $\big(Y_{h_j}^{(j),k}\big)_{k=1,\dots,N_j}$ are independent copies as k varies of $Y_{h_j}^{(j)}$ itself "attached" to $Y_0^{(j)}$ where $(Y_0^{(1)}, \dots, Y_0^{(R)})$ are i.i.d. with the same distribution as Y_0. The size N_j of each simulation at *level* j is of the form $N_j = \lceil q_j N \rceil$, $j = 1, \dots, R$.

If a first order weak expansion error assumption

$$\mathbf{E}\left[Y_h\right] = \mathbf{E}\left[Y_0\right] + c_1 h^{\alpha} + o(h^{\alpha}) \tag{$WE_{\alpha,1}$}$$

is fulfilled for some $\alpha > 0$, then

$$\mathbf{E}\left[\widehat{I}^N_{h,R,q}\right] = \mathbf{E}\left[Y_{h_R}\right] = \mathbf{E}\left[Y_0\right] + c_1\frac{h}{M^{R-1}} + o\big(h/M^{R-1}\big).$$

which dramatically reduces the bias compared to a crude Monte Carlo simulation based on i.i.d. copies of Y_h. At this stage the calibration of the allocation parameters $q_1, \dots, q_R$ across the R levels relies on a strong error convergence rate assumption

$$\forall\, h, h' \in \mathcal{H}, \quad \big\|Y_h - Y_{h'}\big\|_2 \leqslant V_1 |h - h'|^{\beta}, \tag{SE_β}$$

or its variants (see *e.g.* [Pag18] among other references where this conditions are discussed). Thus, it happens that $Y_h - Y_{h'}$ is replaced by a random variable $Y_{h,h'}$ satisfying (SE_β) and such that $\mathbf{E}\left[Y_{h,h'}\right] = \mathbf{E}\left[Y_h - Y_{h'}\right]$. This calibration aims at minimizing the *effort* of the estimator $\widehat{I}^N_{h,q,R}$, that is the product of its variance by its complexity, given a prescribed Root Mean Square Error (RMSE) level $\|\widehat{I}^N_{h,R,q} - Y_0\|_2 \leqslant \varepsilon$.

If a higher order weak error expansion can be established, namely

$$\mathbf{E}\left[Y_h\right] = \mathbf{E}\left[Y_0\right] + \sum_{r=1}^{R} c_r h^{\alpha r} + o(h^{\alpha R}), \tag{$WE_{\alpha,R}$}$$

then there exists *weights* $(\mathbf{w}_j)_{j=1,\ldots,R}$, *only depending on* α, M *and* R such that $\sum_{1\leqslant j\leqslant R} \mathbf{w}_j = 1$ and satisfying

$$\sum_{j=1}^{R} \mathbf{w}_j \mathbf{E}\left[Y_{h_j}\right] = \mathbf{E}\left[Y_0\right] + \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} + o\Big(h^{\alpha R}\Big).$$

These weights, solution to a Vandermonde system (see [LP17]), as well as $\widetilde{\mathbf{w}}_{R+1}$ have closed formulas ($\widetilde{\mathbf{w}}_{R+1} = \sum_{i=1}^{R} \mathbf{w}_i n_i^{-\alpha R}$). This naturally leads to define the *weighted multilevel estimator* (or *Richardson-Romberg multilevel estimator*, ML2R) as

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

where $\mathbf{W}_j^R = \mathbf{w}_j + \cdots + \mathbf{w}_R$, $j = 1, \ldots, R$ and the $Y_{h_j}^{(j),k}$ are as above. One checks that such an estimator "kills" the bias in a much more efficient manner yet since

$$\mathbf{E}\left[\widetilde{I}^N_{h,R,q}\right] = \mathbf{E}\left[Y_0\right] + \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} + o\Big(h^{\alpha R}\Big).$$

Then $\widetilde{I}^N_{h,R,q}$ can be calibrated like the MLMC estimator to minimize its effort for prescribed *RMSE*. For more precise results on the performances of these two families of estimators, we refer to [LP17] or [Gio17] or [Pag18]. But the important fact to be kept in mind is that, as far as nested Monte Carlo simulations are concerned with $f = \mathbf{1}_{[a,+\infty)}$ (see next section for the specification of the r.v. Y_h for this purpose), the β parameter is lower than 1 (see Proposition 5.2) so that, as a consequence, the ML2R estimator $\widetilde{I}^N_{h,R,q}$ behaves "almost" like an unbiased estimator, for which the cost is known to be $K\varepsilon^{-2}$, $K > 0$ constant. More precisely, if ($WE_{\alpha,R}$) holds for every *depth* $R \geqslant 1$ and $\lim_{R\to\infty} |c_R|^{\frac{1}{R}} = \widetilde{c}_\infty \in (0, +\infty)$, then

$$\text{Cost}\left(\widetilde{I}^{N(\varepsilon)}_{h(\varepsilon),R(\varepsilon),q(\varepsilon)}\right) \preceq K_{\alpha,\beta,M}\,\varepsilon^{-2} \cdot e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}},$$

where we recall that $f(\varepsilon) \preceq g(\varepsilon)$ if and only if $\limsup_{\varepsilon\to 0} g(\varepsilon)/f(\varepsilon) \leq 1$, $K_{\alpha,\beta,M} > 0$ is constant and we highlight that $e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}} = o\left(\varepsilon^{-\eta}\right)$ for all $\eta > 0$. Note that some numerical experiments carried out in [LP17] and in [Gio17] confirm the fact that weighted multilevel ML2R simulations outperform regular MLMC estimator.

The paper is organized as follows. In Section 2 we give the description of the nested framework. In Section 3 we give some useful results which will be valid in both frameworks, both f smooth and not. Section 4 is devoted to the smooth case, with a particular attention to the antithetic approach, and in Section 5 we treat the non smooth case and we give a new result concerning the weak error.

2 Nested Monte Carlo simulation

The purpose of the so-called *nested* Monte Carlo method is to compute by simulation nested expectations of the form

$$\mathbf{E}\left[f\left(\mathbf{E}\left[\Xi|Y\right]\right)\right],$$

where (Ξ, Y) is an $\mathbf{R}\times\mathbf{R}^d$-valued couple of random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ satisfying $\Xi\in L^2$ and $f:\mathbf{R}\to\mathbf{R}$ is a specified function such that $f(\mathbf{E}\left[\Xi|Y\right])\in L^2$.

We assume that there exist a Borel function $F:\mathbf{R}^d\times\mathbf{R}^q\to\mathbf{R}$ and a random vector $Z:(\Omega,\mathcal{A})\to\mathbf{R}^q$ independent of Y such that

$$\Xi = F(Y,Z).$$

Let us introduce the Borel function $\phi_0:\mathbf{R}^d\to\mathbf{R}$ defined by $\phi_0(y)=\mathbf{E}\left[F(y,Z)\right]$ so that one may set $\mathbf{E}\left[F(Y,Z)|Y\right]=\phi_0(Y)$. Then one has the following representation

$$\mathbf{E}\left[\Xi|Y\right]=\phi_0(Y)=\int_{\mathbf{R}^q}F(Y,z)\,\mathbf{P}_Z(\mathrm{d}z).$$

To comply with the multilevel framework, we set $K_0\in\mathbf{N}^*$ and $\mathcal{H}=\left\{1/K,\ K\in K_0\mathbf{N}^*\right\}$,

$$X_0:=\mathbf{E}\left[\Xi|Y\right],\quad X_h:=\frac{1}{K}\sum_{k=1}^{K}F(Y,Z_k)\ \text{ with }\ h=\frac{1}{K}\in\mathcal{H},$$

where $(Z_k)_{k\geqslant 1}$ is an i.i.d. sequence of random vectors with the same distribution as Z, defined on $(\Omega,\mathcal{A},\mathbf{P})$ and independent of Y (up to an enlargement of the probability space if necessary) and

$$Y_0:=f(X_0),\quad Y_h:=f(X_h).$$

To prove that the nested Monte Carlo estimator satisfies the bias error expansion ($WE_{\alpha,R}$) and the strong approximation error (SE_β), we introduce the random functions, $\forall y\in\mathbf{R}^d$,

$$D(y)=F(y,Z)-\mathbf{E}\left[F(y,Z)\right],\tag{3}$$

$$E_h(y)=\frac{1}{K}\sum_{k=1}^{K}\Big(F(y,Z_k)-\mathbf{E}\left[F(y,Z)\right]\Big)=\frac{1}{K}\sum_{k=1}^{K}F(y,Z_k)-\phi_0(y).\tag{4}$$

Note that $E_h(y)$ is the statistical error of the inner Monte Carlo estimator, which can be rewrited as $E_h(y)=\dfrac{1}{K}\sum_{k=1}^{K}D(y)^{(k)}$ where $(D(y)^{(k)})_{k\geqslant 1}$ is a sequence of $i.i.d.$ copies of $D(y)$, and that $E_h(Y)=X_h-X_0$.

We distinguish between two main frameworks, depending on whether or not f is smooth, a classical example of non-smoothness being $f=\mathbf{1}_{(a,b)}$ (see [DL09]). When the function f is smooth enough, say $f \in \mathcal{C}^{1+\rho}(\mathbf{R},\mathbf{R})$ with $\rho \in (0,1]$, a variant of the former Multilevel *nested* estimator has been proposed in [BHR15], [Haj12] and [CL12] (see also [Gil15]) to improve the rate of strong convergence in order to attain the asymptotically unbiased setting, namely (SE_β) with $\beta > 1$. A root $M \geqslant 2$ being given, the idea is to replace in the successive refined levels of the MLMC and ML2R estimators (see (1) and (2)) the difference $Y_{\frac{h}{M}} - Y_h$ (where $h = \frac{1}{K}$, $K \in K_0\mathbf{N}^*$) by an antithetic type as follows

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

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

Before getting into the smooth and non smooth case, we give some useful results that will be valid in both frameworks and will be used to establish the higher order of weak error expansion.

3 Useful results

Following Comtet [Com74], we introduce the partial Bell polynomials $\mathrm{B}_{n,k}$ for $n \geqslant 1$ and $k = 1, \ldots, n$ defined by

$$\mathrm{B}_{n,k}(x_1, \ldots, x_{n-k+1}) = \sum \frac{n!}{\ell_1! \cdots \ell_{n-k+1}!}\Big(\frac{x_1}{1!}\Big)^{\ell_1} \cdots \Big(\frac{x_{n-k+1}}{(n-k+1)!}\Big)^{\ell_{n-k+1}} \tag{5}$$

where the summation takes place over all integers $\ell_1, \ldots, \ell_n \geqslant 0$, such that $\ell_1 + 2\ell_2 + \cdots + (n-k+1)\ell_{n-k+1} = n$ and $\ell_1 + \cdots + \ell_{n-k+1} = k$. Note that $\deg(B_{n,k}) = k$. The complete Bell polynomials B_n are defined by

$$\mathrm{B}_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} \mathrm{B}_{n,k}(x_1, \ldots, x_{n-k+1}).$$

The first statement is a formal Taylor expansion with integral remainder of $\mathbf{E}\left[g(X_h)\right]$ around $\mathbf{E}\left[g(X_0)\right]$, with $g : \mathbf{R} \to \mathbf{R}$ a test function.

Lemma 3.1 (Taylor expansion)**.** *Let $R \geqslant 0$ and let $g : \mathbf{R} \to \mathbf{C}$ be a $2R+1$ times differentiable function.*

Assume $\Xi = F(Y, Z) \in L^{2R+1}$ and let $\kappa_j(\xi)$ be the j–th cumulant (a.k.a. semi-invariant) of a random variable ξ. We set $\kappa_{j,y} := \kappa_j(D(y))$ with $D(y) = F(y, Z) - \mathbf{E}\left[F(y, Z)\right]$ for $y \in \mathbf{R}^d$ and $j \in \big\{1, \ldots, R\big\}$. Let $(\mathrm{B}_{n,k})_{1 \leqslant k \leqslant n}$ be the partial Bell polynomials *defined by (5). We then define for $r \in \mathbf{N}$, $r + 1 \leqslant n \leqslant 2r$ and every $y \in \mathbf{R}^d$,*

$$b_{r,n-r}(y) = \mathrm{B}_{r,n-r}\left(\frac{\kappa_{2,y}}{2}, \ldots, \frac{\kappa_{2r-n+2,y}}{2r-n+2}\right).$$

Then

$$\forall h \in \mathcal{H}, \quad \mathbf{E}\left[g(X_h)\right] = \mathbf{E}\left[g(X_0)\right] + \sum_{r=1}^{2R-1} c(r, (2r+1) \wedge 2R)h^r + \mathcal{R}_{2R+1}, \tag{6}$$

with

$$c(r,k) = \frac{1}{r!}\sum_{\ell=r+1}^{k} \mathbf{E}\left[g^{(\ell)}(X_0) b_{r,\ell-r}(Y)\right],\ 1 \leqslant r < k \leqslant 2r+1, \tag{7}$$

and

$$\mathcal{R}_{2R+1} = \frac{1}{(2R)!}\mathbf{E}\left[\int_0^{X_h - X_0} g^{(2R+1)}(t+X_0)(X_h - X_0 - t)^{2R}\mathrm{d}t\right]. \tag{8}$$

Proof. The case $R=0$ is trivial, since it is a direct application of the fundamental theorem of calculus.

Let $R \geqslant 1$ be an integer. The Taylor formula at order $2R$ applied to g at $\phi_0(y)$ reads

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(y, Z_k)\right)\right] = g(\phi_0(y)) + \sum_{n=1}^{2R}\frac{g^{(n)}(\phi_0(y))}{n!}\mathbf{E}\left[(E_h(y))^n\right] + \mathcal{R}_{2R+1}(y), \tag{9}$$

where $\mathcal{R}_{2R+1}(y) = \frac{1}{(2R)!}\mathbf{E}\left[\int_0^{E_h(y)} g^{(2R+1)}(t+\phi_0(y))(E_h(y)-t)^{2R}\mathrm{d}t\right]$.

The Bell polynomials allow us to explicitly compute the moments $\mathbf{E}\big[(E_h(y))^n\big]$, $n=1,\dots,R$ of $E_h(y)$ as follows. Let $\kappa_{j,y} = \kappa_j(D(y))$, $j=1,\dots,R$, $y \in \mathbf{R}^d$. Additivity and homogeneity of cumulants give

$$\forall j = 1,\dots,2R-1, \quad \kappa_j\big(E_h(y)\big) = h^{j-1}\kappa_{j,y}.$$

Moments of $E_h(y)$ can be expressed in terms of cumulants using complete Bell polynomials (see [Com74] p.160 Equation(2)) as:

$$\mathbf{E}\big[(E_h(y))^n\big] = \mathrm{B}_n\Big(\kappa_1\big(E_h(y)\big),\dots,\kappa_n\big(E_h(y)\big)\Big).$$

First note that $\kappa_{1,y} = 0$ so that $\kappa_1\big(E_h(y)\big) = 0$. Moreover, it follows from the definition (5) that $\mathrm{B}_{n,k}$ is k–homogeneous, consequently

$$\mathbf{E}\big[(E_h(y))^n\big] = h^n \sum_{k=1}^{n} h^{-k}\,\mathrm{B}_{n,k}\big(0,\kappa_{2,y},\dots,\kappa_{n-k+1,y}\big).$$

We again derive from (5) that $\mathrm{B}_{n,n}(0) = 0$, hence the last term in the above sum is null. In particular the sum in (9) starts from $n=2$. Note now that

$$\mathrm{B}_{n,k}\left(0,\kappa_{2,y},\dots,\kappa_{n-k+1,y}\right) = \begin{cases} \frac{n!}{(n-k)!} b_{n-k,k}(y) & \text{if } 1 \leqslant k \leqslant \lceil n/2 \rceil, \\ 0 & \text{if } k > \lceil n/2 \rceil, \end{cases}$$

with $b_{n-k,k}(y) = \mathrm{B}_{n-k,k}\left(\frac{\kappa_{2,y}}{2},\dots,\frac{\kappa_{n-2k+2,y}}{n-2k+2}\right)$ which implies that

$$\mathbf{E}\big[(E_h(y))^n\big] = h^n \sum_{k=1}^{\lceil n/2 \rceil} h^{-k}\frac{n!}{(n-k)!} b_{n-k,k}(y). \tag{10}$$

Plugging (10) in (9) gives, since the sum starts at $n = 2$ as mentioned above,

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(y, Z_k)\right)\right] = g(\phi_0(y)) + \sum_{n=2}^{2R} g^{(n)}(\phi_0(y)) \sum_{k=1}^{\lceil n/2\rceil} \frac{h^{n-k}}{(n-k)!} b_{n-k,k}(y) + \mathcal{R}_{2R+1}(y).$$

Setting $r = n - k$ in the above expression, noting that $\lceil n/2\rceil + \lfloor n/2\rfloor = n$ and that $\lfloor n/2\rfloor \leqslant r$ if and only if $n \leq 2r + 1$, on derives by interchanging the sums that

$$\mathbf{E}\left[g\left(\frac{1}{K}\sum_{k=1}^{K} F(y, Z_k)\right)\right] = g(\phi_0(y)) + \sum_{r=1}^{2R-1} \frac{h^r}{r!}\left(\sum_{n=r+1}^{(2r+1)\wedge 2R} g^{(n)}(\phi_0(y)) b_{r,n-r}(y)\right) + \mathcal{R}_{2R+1}(y).$$

We conclude by integrating with respect to $\mathbf{P}_Y(\mathrm{d}y)$. □

Taking advantage of this expansion we will derive two results. First a bias error expansion for smooth enough payoff functions, in which no regularity is required on the law of (X_0, X_h) (see Subsection 4.1). Conversely a second result will be established relying on the regularity of the distribution of (X_0, X_h) when the payoff function is not smooth (see Subsection 5.1).

As concerns the strong error, elementary computations show that, if $\Xi \in L^2$, then

$$\left\|X_h - X_0\right\|_2^2 = \frac{1}{K}\int \mathbf{P}_Y(\mathrm{d}y)\mathrm{var}\big(F(y, Z)\big) = h\mathbf{E}\left[(F(Y, Z) - \phi_0(Y))^2\right] \leq h\,\mathrm{var}\big(F(Y, Z)\big), \tag{11}$$

since $\phi_0(Y) = \mathbf{E}\left[F(Y, Z)|Y\right]$. To prove ($SE_\beta$) we extend this result to $\left\|X_h - X_{h'}\right\|_p$ when $\Xi \in L^p$, $p > 1$, as described in Lemma 3.2.

Note that from now on we give the results for a generic $p > 1$ instead of $p = 2$, because this wider assumption can be useful to establish a condition of uniform integrability needed to prove a Central Limit Theorem (and strong law of large numbers) for Multilevel Monte Carlo estimators, see Lemma 5.2 in [GLP17].

The proof of Lemma 3.2 relies on the Marcinkiewicz-Zygmund inequality that we recall for clarity. If $(\xi_n)_{n\geqslant 1}$ is a sequence of centered independent random variables such that $\mathbf{E}\left[\left|\xi_n\right|^p\right] < +\infty$, $1 < p < +\infty$, then

$$\left\|\sum_{k=1}^{K}\xi_k\right\|_p \leqslant (B_p)^{\frac{1}{p}}\left\|\sum_{k=1}^{K}\xi_k^2\right\|_{\frac{p}{2}}^{\frac{1}{2}}, \tag{12}$$

where $B_p = \frac{18p^{\frac{3}{2}}}{(p-1)^{\frac{1}{2}}}$ (see [Shi96] p.499). If moreover $(\xi_n)_{n\geqslant 1}$ are identically distributed we have

$$\left\|\sum_{k=1}^{K}\xi_k\right\|_p \leqslant (B_p)^{\frac{1}{p}}\sqrt{K}\left\|\xi_1\right\|_p \tag{13}$$

We make an intensive use of this inequality in Section 4.

Lemma 3.2. *Assume $\Xi \in L^p$, $p > 1$. Then, for every h, $h' \in \mathcal{H}$,*

$$\left\|X_h - X_{h'}\right\|_p \leqslant 2B_p\left\|\Xi - \mathbf{E}\left[\Xi|Y\right]\right\|_p |h - h'|^{\frac{1}{2}}. \tag{14}$$

Proof. Assume first that $h' \leqslant h$, and set $K = \frac{1}{h}$, $K' = \frac{1}{h'} \geqslant K$. First note that by Fubini's theorem

$$\left\|X_h - X_{h'}\right\|_p^p = \int_{\mathbf{R}^d} \mathbf{P}_{_Y}(\mathrm{d}y)\mathbf{E}\left[|E_h(y) - E_{h'}(y)|^p\right].$$

Setting $\widetilde{F}(y,z) = F(y,z) - \phi_0(y)$, we write

$$E_h(y) - E_{h'}(y) = \left(\frac{1}{K} - \frac{1}{K'}\right)\sum_{k=1}^{K} \widetilde{F}(y, Z_k) + \frac{1}{K'}\sum_{k=K+1}^{K'} \widetilde{F}(y, Z_k).$$

Then, for every $y \in \mathbf{R}^d$, it follows from Minkowski's Inequality,

$$\left\|E_h(y) - E_{h'}(y)\right\|_p \leqslant |h - h'| \left\|\sum_{k=1}^{K} \widetilde{F}(y, Z_k)\right\|_p + h' \left\|\sum_{k=K+1}^{K'} \widetilde{F}(y, Z_k)\right\|_p.$$

Applying Marcinkiewicz-Zygmund Inequality to both terms on the right hand side of the above inequality yields

$$\begin{aligned}\left\|E_h(y) - E_{h'}(y)\right\|_p &\leqslant |h - h'| B_p \left\|\sum_{k=1}^{K} \widetilde{F}(y, Z_k)^2\right\|_{\frac{p}{2}}^{\frac{1}{2}} + h' B_p \left\|\sum_{k=K+1}^{K'} \widetilde{F}(y, Z_k)^2\right\|_{\frac{p}{2}}, \\ &\leqslant |h - h'| B_p K^{\frac{1}{2}} \left\|\widetilde{F}(y, Z)\right\|_p + h' B_p (K' - K)^{\frac{1}{2}} \left\|\widetilde{F}(y, Z)\right\|_p.\end{aligned}$$

Finally, for every $y \in \mathbf{R}^d$,

$$\begin{aligned}\left\|E_h(y) - E_{h'}(y)\right\|_p &\leqslant B_p \left\|\widetilde{F}(y, Z)\right\|_p \left((h - h')\frac{1}{\sqrt{h}} + h'\left(\frac{1}{h'} - \frac{1}{h}\right)^{\frac{1}{2}}\right) \\ &= B_p \left\|\widetilde{F}(y, Z)\right\|_p (h - h')^{\frac{1}{2}} \left(\left(1 - \frac{h'}{h}\right)^{\frac{1}{2}} + \left(\frac{h'}{h}\right)^{\frac{1}{2}}\right) \\ &\leqslant 2B_p \left\|\widetilde{F}(y, Z)\right\|_p (h - h')^{\frac{1}{2}}.\end{aligned}$$

Plugging this bound in the above equality yields, owing to Minkowski's Inequality and Jensen's Inequality for conditional expectations, the announced result

$$\begin{aligned}\left\|X_h - X_{h'}\right\|_p^p &\leqslant (2B_p)^p \int_{\mathbf{R}^d} \mathbf{P}_{_Y}(\mathrm{d}y) \left\|\widetilde{F}(y, Z)\right\|_p^p (h - h')^{\frac{p}{2}} \\ &= (2B_p)^p \left\|\Xi - \mathbf{E}\left[\Xi | Y\right]\right\|_p^p (h - h')^{\frac{p}{2}}.\end{aligned}$$

$\square$

4 Smooth payoff function

We first focus on the smooth case, where we give a bias error expansion and a strong convergence rate when the payoff function f is smooth. This result beyond its direct application will be an important step when dealing with indicator functions.

4.1 Weak error

The bias error expansion of the nested Monte Carlo estimator when f is smooth is a consequence of Lemma 3.1, as we emphasized in the proof of the following Proposition.

Proposition 4.1 (Bias error (I): smooth functions)**.** *Let $R \in \mathbf{N}^*$ and let $f : \mathbf{R} \to \mathbf{R}$ be a $2R+1$ times differentiable payoff function with bounded derivatives $f^{(k)}$, $k = R+1, \ldots, 2R+1$. Assume $X \in L^{2R+1}$. Then there exists $c_1, \ldots, c_R$ such that*

$$\forall h \in \mathcal{H}, \quad \mathbf{E}\left[f(X_h)\right] = \mathbf{E}\left[f(X_0)\right] + \sum_{r=1}^{R} c_r h^r + \mathcal{O}(h^{R+1/2}). \tag{15}$$

Proof. Applying Lemma 3.1 with the function $g = f$ we get, for every $h \in \mathcal{H}$,

$$\mathbf{E}\left[f(X_h)\right] = \mathbf{E}\left[f(X_0)\right] + \sum_{r=1}^{R-1} c(r, 2r+1)h^r + c(R, 2R)h^R + \sum_{r=R+1}^{2R-1} c(r, 2R)h^r + \mathcal{R}_{2R+1}, \tag{16}$$

with $c(r, k)$ defined in (7) and $\mathcal{R}_{2R+1}$ in (8). Establishing the proposition amounts to proving that the remainder term $\mathcal{R}_{2R+1}$ is well controlled. Using that $f^{(2R+1)}$ is bounded, we have

$$\left|\mathcal{R}_{2R+1}\right| \leqslant \frac{\left\|f^{(2R+1)}\right\|_\infty}{(2R+1)!} \mathbf{E}\left[\left|X_h - X_0\right|^{2R+1}\right].$$

Using successively the Marcinkiewicz-Zygmund Inequality and the Minkowski Inequality for the $L^{R+\frac{1}{2}}(\mathbf{P})$-norm, we get, keeping in mind that $h = \frac{1}{K}$,

$$\begin{aligned}
\mathbf{E}\left[\left|E_h(y)\right|^{2R+1}\right] &\leqslant (B_{2R+1})^{2R+1} h^{2R+1} \mathbf{E}\left[\left|\sum_{k=1}^{K} (F(y, Z_k) - \phi_0(y))^2\right|^{R+1/2}\right] \\
&\leqslant (B_{2R+1})^{2R+1} h^{R+1/2} \mathbf{E}\left[\left|F(y, Z) - \phi_0(y)\right|^{2R+1}\right]
\end{aligned}$$

Integrating with respect to $\mathbf{P}_Y$ finally yields

$$\begin{aligned}
\mathbf{E}\left[\left|X_h - X_0\right|^{2R+1}\right] &\leqslant (B_{2R+1})^{2R+1} h^{R+1/2} \mathbf{E}\left[\left|F(Y, Z) - \phi_0(Y)\right|^{2R+1}\right] \\
&\leqslant 2^{R+\frac{1}{2}} (B_{2R+1})^{2R+1} h^{R+1/2} \mathbf{E}\left[|\Xi|^{2R+1} + |\mathbf{E}\left[\Xi|Y\right]|^{2R+1}\right] \\
&\leqslant 2^{R+\frac{3}{2}} (B_{2R+1})^{2R+1} h^{R+1/2} \mathbf{E}\left[|\Xi|^{2R+1}\right]
\end{aligned} \tag{17}$$

so that $\left|\mathcal{R}_{2R+1}\right| = \mathcal{O}(h^{R+1/2})$. □

4.2 Strong convergence rate

If we assume that f is Lipschitz continuous, Lemma 3.2 straightforwardly shows that the standard *nested* Monte Carlo satisfies a strong convergence at a rate h^β with $\beta = 1$. More precisely, if $\Xi \in L^2$, we have

$$\left\|Y_h - Y_{h'}\right\|_2 \leqslant 2B_2[f]_{\text{Lip}} \left\|\Xi - \mathbf{E}\left[\Xi|Y\right]\right\|_2 \left|h - h'\right|^{\frac{1}{2}},$$

where $[f]_{\text{Lip}}$ denotes the Lipschitz coefficient of f.

When asking for more smoothness, more precisely that f' is ρ–Hölder, we can build an antithetic version of the *nested* Monte Carlo which attains a strong convergence at a rate h^{β} with $\beta > 1$. As we saw, this corresponds to the optimal unbiased setting in terms of minimization of the computational cost. This antithetic multilevel estimator is obtained by replacing each difference $Y_{\frac{h}{M}} - Y_h$, $h \in h_1, \ldots, h_{R-1}$, in the MLMC (1) and ML2R (2) estimators by the following random variable

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

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

$$\bar{X}_{K,m} = \frac{1}{K}\sum_{k=1}^{K} F(Y, Z_{K(m-1)+k}) \quad \text{and} \quad \bar{X}_{MK} = \frac{1}{M}\sum_{m=1}^{M} \bar{X}_{K,m} = \frac{1}{MK}\sum_{k=1}^{MK} F(Y, Z_k),$$

so that the nested antithetic MLMC estimator (1) then reads

$$\widehat{I}^{N}_{h,R,q} = \frac{1}{N_1}\sum_{i=1}^{N_1} f(\bar{X}^{(i)}_{K,1}) + \sum_{j=2}^{R}\frac{1}{N_j}\sum_{i=1}^{N_j}\left(f(\bar{X}^{(i)}_{MK}) - \frac{1}{M}\sum_{m=1}^{M} f(\bar{X}^{(i)}_{K,m})\right),$$

with $(\bar{X}^{(i)}_{K,m})_{i\geq 1}$ independent copies of $\bar{X}_{K,m}$, and similarly for the ML2R estimator (2) with the weights $(W_j)_{2\leqslant j\leqslant R}$.

Proposition 4.2. *Let $p > 1$ and $0 < \rho \leqslant 1$. Assume $\Xi \in L^{p(1+\rho)}$ and f' ρ–Hölder, i.e.*

$$\forall x, y \in \mathbf{R}, \quad \big|f'(x) - f'(y)\big| \leq \big[f'\big]_{\rho}\,|x-y|^{\rho}. \tag{18}$$

Then $Y_{h,\frac{h}{M}}$ satisfies a strong approximation error control similar as (SE_{β}) with $\beta = 1+\rho > 1$. More precisely, we prove that there exists $\widetilde{V}_1 > 0$ depending only on $p, \rho, [f']_{\rho}$ and M such that

$$\left\|Y_{h,\frac{h}{M}}\right\|_p \leqslant \widetilde{V}_1\left(h - \frac{h}{M}\right)^{\frac{1+\rho}{2}}. \tag{19}$$

Proof. Owing to Taylor's formula, for all $m = 1, \ldots, M$, there exists x_m in the geometric segment $\big(\bar{X}_{K,m}, \bar{X}_{MK}\big)$ such that

$$f(\bar{X}_{K,m}) = f(\bar{X}_{MK}) + f'(\bar{X}_{MK})(\bar{X}_{K,m} - \bar{X}_{MK}) + \big(f'(x_m) - f'(\bar{X}_{MK})\big)\,(\bar{X}_{K,m} - \bar{X}_{MK}).$$

Hence, using the definition of $\bar{X}_{MK}$,

$$\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) = f(\bar{X}_{MK}) + \frac{1}{M}\sum_{m=1}^{M}\big(f'(x_m) - f'(\bar{X}_{MK})\big)\,(\bar{X}_{K,m} - \bar{X}_{MK}). \tag{20}$$

We aim at computing $\left\|Y_{h,\frac{h}{M}}\right\|_p = \left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p$. Owing to the decomposition (20), to Minkowski's Inequality and to the ρ–Hölder assumption (18) on f', we get

$$\left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p = \left\|\frac{1}{M}\sum_{m=1}^{M} \big(f'(x_m) - f'(\bar{X}_{MK})\big)\left(\bar{X}_{K,m} - \bar{X}_{MK}\right)\right\|_p$$
$$\leq [f']_\rho \frac{1}{M}\sum_{m=1}^{M}\left\|X_{K,m} - \bar{X}_{MK}|^{1+\rho}\right\|_p. \tag{21}$$

We first notice by an exchangeability argument that the variables $(\bar{X}_{K,m} - \bar{X}_{MK})_{m=1\ldots,M}$ are identically distributed with $\bar{X}_{K,m} - \bar{X}_{MK} \sim \bar{X}_{K,1} - \bar{X}_{MK}$. Moreover we write

$$\bar{X}_{K,1} - \bar{X}_{MK} = \bar{X}_{K,1} - \frac{1}{M}\sum_{m=1}^{M}\bar{X}_{K,m} = \frac{1}{M}\sum_{m=1}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m}) = \frac{1}{M}\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m}).$$

Hence, we get

$$\left\|\frac{1}{M}\sum_{m=1}^{M} f(\bar{X}_{K,m}) - f(\bar{X}_{MK})\right\|_p \leq [f']_\rho \frac{1}{M^{1+\rho}}\left\|\left|\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{1+\rho}\right\|_p. \tag{22}$$

Owing to the independence of Y and $(Z_k)_{k\geq 1}$, we may write

$$\begin{aligned}
&\mathbf{E}\left[\left|\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{(1+\rho)p}\right] \\
&= \int \mathbf{P}_Y(\mathrm{d}y)\mathbf{E}\left[\left|\sum_{m=2}^{M}\left(\frac{1}{K}\sum_{k=1}^{K}F(y,Z_k) - \frac{1}{K}\sum_{k=1}^{K}F(y,Z_{K(m-1)+k})\right)\right|^{(1+\rho)p}\right] \\
&= \int \mathbf{P}_Y(\mathrm{d}y)\frac{1}{K^{(1+\rho)p}}\mathbf{E}\left[\left|\sum_{k=1}^{K}\left((M-1)F(y,Z_k) - \sum_{m=2}^{M}F(y,Z_{K(m-1)+k})\right)\right|^{(1+\rho)p}\right].
\end{aligned}$$

We notice that, for each fixed $y \in \mathbf{R}^d$, the random variables $\xi_k = (M-1)F(Z_k,y) - \sum_{m=2}^{M}F(y,Z_{K(m-1)+k})$, $k \geqslant 1$, are centered and *i.i.d.*. Moreover $(1+\rho)p > 1$ hence, owing to Marcinkiewicz-Zygmund inequality (13), we have

$$\begin{aligned}
&\mathbf{E}\left[\left|\sum_{m=2}^{M}(\bar{X}_{K,1} - \bar{X}_{K,m})\right|^{(1+\rho)p}\right] \\
&\qquad\leq B_{(1+\rho)p}\frac{1}{K^{\frac{(1+\rho)p}{2}}}\int \mathbf{P}_Y(\mathrm{d}y)\mathbf{E}\left[\left|(M-1)F(y,Z_1) - \sum_{m=2}^{M}F(y,Z_{K(m-1)+1})\right|^{(1+\rho)p}\right].
\end{aligned}$$

Applying twice Minkowski's Inequality yields

$$\mathbf{E}\left[\left|\sum_{m=2}^{M}(\bar{X}_{K,1}-\bar{X}_{K,m})\right|^{(1+\rho)p}\right] \leq C_{p,\rho}\frac{1}{K^{\frac{(1+\rho)p}{2}}}(M-1)^{(1+\rho)p}2^{(1+\rho)p}\mathbf{E}\left[|F(Y,Z_1)|^{(1+\rho)p}\right].$$

Plugging this in (22) we get

$$\left\|\frac{1}{M}\sum_{m=1}^{M}f(\bar{X}_{K,m})-f(\bar{x})\right\|_p^p \leq \widetilde{V}_1\left(1-\frac{1}{M}\right)^{\frac{p(1+\rho)}{2}}\frac{1}{K^{\frac{p(1+\rho)}{2}}},$$

with $\widetilde{V}_1=[f']_\rho^p C_{p,\rho}2^{p(1+\rho)}\left(1-\frac{1}{M}\right)^{\frac{(1+\rho)p}{2}}\|\Xi\|_{(1+\rho)p}^{(1+\rho)p}$, and (19) is proved. □

If we replace the ρ–Hölder assumption on f' by a weaker assumption f' locally ρ–Hölder, *i.e.*

$$\forall x,y\in\mathbf{R},\quad |f'(x)-f'(y)|\leq C|x-y|^\rho\left(1+|x|^q+|y|^q\right),$$

a strong convergence assumption with $\beta=1+\rho>1$ similar to (19) can still be proved. Since $|x_m|^q\leq\max(|\bar{X}_{K,m}|^q,|\bar{X}_{MK}|^q)\leq|\bar{X}_{K,m}|^q+|\bar{X}_{MK}|^q$, Inequality (21) must be replaced by

$$\begin{aligned}\left\|\frac{1}{M}\sum_{m=1}^{M}f(\bar{X}_{K,m})-f(\bar{X}_{MK})\right\|_p &= \left\|\frac{1}{M}\sum_{m=1}^{M}\left(f'(x_m)-f'(\bar{X}_{MK})\right)(\bar{X}_{K,m}-\bar{X}_{MK})\right\|_p\\ &\leq [f']_\rho\frac{1}{M}\sum_{m=1}^{M}\left\||\bar{X}_{K,m}-\bar{X}_{MK}|^{1+\rho}(1+|\bar{X}_{K,m}|^q+2|\bar{X}_{MK}|^q)\right\|_p.\end{aligned}\tag{23}$$

Owing to Hölder's Inequality with $r,s>1$ such that $\frac{1}{r}+\frac{1}{s}=1$ and Minkowski's Inequality, we get

$$\begin{aligned}&\left\||\bar{X}_{K,m}-\bar{X}_{MK}|^{1+\rho}(1+|\bar{X}_{K,m}|^q+2|\bar{X}_{MK}|^q)\right\|_p\\ &\qquad\leq\left\||\bar{X}_{K,m}-\bar{X}_{MK}|^{1+\rho}\right\|_{pr}\left(1+\left\||\bar{X}_{K,m}|^q\right\|_{ps}+2\left\||\bar{X}_{MK}|^q\right\|_{ps}\right).\end{aligned}$$

Since the variables $(\bar{X}_{K,m})_{m=1,\ldots,M}$ are identically distributed, Inequality (23) yields

$$\begin{aligned}&\left\|\frac{1}{M}\sum_{m=1}^{M}f(\bar{X}_{K,m})-f(\bar{X}_{MK})\right\|_p\\ &\qquad\leq[f']_\rho\left\||\bar{X}_{K,1}-\bar{X}_{MK}|^{1+\rho}\right\|_{pr}\left(1+\left\||\bar{X}_{K,m}|^q\right\|_{ps}+2\left\||\bar{X}_{MK}|^q\right\|_{ps}\right).\end{aligned}$$

The analysis of the term $\left\||\bar{X}_{K,1}-\bar{X}_{MK}|^{1+\rho}\right\|_{pr}$ does not change, except for the condition $\Xi\in L^{(1+\rho)pr}$. Under the assumption $\Xi\in L^{qps}$, the term $\left\||\bar{X}_{K,m}|^q\right\|_{ps}+2\left\||\bar{X}_{MK}|^q\right\|_{ps}$ is bounded, since

$$\left\||X_{K,1}|^q\right\|_{ps}=\left\|\left|\frac{1}{K}\sum_{k=1}^{K}F(Y,Z_k)\right|^q\right\|_{ps}\leq\|\Xi\|_{qps\vee1}^q.$$

Keeping in mind that $r = s/(s-1)$, the optimal choice for s which minimizes both $(1+\rho)pr$ and qps is given by $s = (1+\rho+q)/q$ (hence $r = (1+\rho+q)/(1+\rho)$). This leads to the additional condition $\Xi \in L^{p(1+\rho+q)}$. In conclusion, if f' is locally ρ–Hölder, under the assumption $\Xi \in L^{p(1+\rho+q)}$, $Y_{h,\frac{h}{M}}$ satisfies the L^p version of the strong convergence assumption with $\beta = 1+\rho > 1$, similarly to (19).

5 Indicator function and smooth density

There are many situations where we need to consider non smooth payoff functions of the type $f = \mathbf{1}_{\{g(\mathbf{E}[\Xi|Y])\in I\}}$, with $g : \mathbf{R} \to \mathbf{R}$ and $I \subset \mathbf{R}$ interval. Among them we can cite the computation of loss thresholds, *i.e.* when we search, a threshold $q \in \mathbf{R}$ being fixed, for the corresponding $\alpha_q \in [0,1]$ such that

$$1-\alpha_q = \mathbf{P}(g(\mathbf{E}\left[\Xi|Y\right]) \geq q) = \mathbf{E}\left[\mathbf{1}_{\{g(\mathbf{E}[\Xi|Y])\geq q\}}\right],$$

or the inverse problem, which consists in computing the quantile q_α such that for a fixed $\alpha \in [0,1]$,

$$1-\alpha = \mathbf{P}(g(\mathbf{E}\left[\Xi|Y\right]) \geq q_\alpha) = \mathbf{E}\left[\mathbf{1}_{\{g(\mathbf{E}[\Xi|Y])\geq q_\alpha\}}\right].$$

Another situation of interest is the approximation of density functions (see the seminal paper of Bally and Talay [BT96a] and [BT96b], treating the law of the Euler scheme for distributions).

The payoff function f being non smooth, the regularity assumptions on f that we needed to prove the weak and the strong convergence of the estimator in the smooth case, will be replaced by some regularity assumptions on the density functions, as we detail in the next two Subsections.

5.1 Weak error

We recall the notation that $X_0 = \mathbf{E}\left[\Xi|Y\right]$ and $X_h = \frac{1}{K}\sum_{k=1}^{K} F(Y, Z_k)$ with $h = \frac{1}{K} \in \mathcal{H}$ and we introduce the notation

$$\Delta_h = X_h - X_0.$$

The following result on the weak error derives from Lemma 3.1 and gives a bias error expansion relying on the density of the joint distribution of (X_0, Δ_h) and of (X_0, Y). More precisely, assume that (X_0, Δ_h) is a random vector with smooth density with respect to the Lebesgue measure on $\mathbf{R}^2$. Let f_{X_0} be the density of X_0, let $f_{X_0,Y}$ be the density of (X_0, Y) and let f_{X_0,Δ_h} be the density of (X_0, Δ_h). Moreover let $F_{X_h}(x)$ and $F_{X_0}(x)$ be the cumulative distribution functions of Δ_h and X_0.

Proposition 5.1 (Bias error (II): smooth density)**.** *(a) Let $R \geqslant 0$. Assume that the partial derivatives $\partial_x^{(\ell)} f_{X_0,Y}(x,y)$ exist for $\ell = 1,\ldots,2R$, that the partial derivatives $\partial_x^{(\ell)} f_{X_0,\Delta_h}(x,y)$ exist for $\ell = 1,\ldots,2R+1$ and that $\partial_x^{(2R+1)} f_{X_0,\Delta_h}(x,y)$ is continuous. Assume that $\Xi \in L^{2R+1}$.*

Let

$$P_r(x) = \frac{1}{f_{X_0}(x)} \sum_{\ell=r+1}^{(2r+1)\wedge 2R} (-1)^\ell \int_{\mathbf{R}} b_{r,\ell-r}(y)\partial_x^{(\ell)} f_{X_0,Y}(x,y)\mathrm{d}y.$$

(*a*) *If* $\sup_{h\in\mathcal{H},x,v\in\mathbf{R}}\left|\partial_x^{(2R+1)}f_{X_0|\Delta_h=v}(x)\right|<+\infty$, *then*

$$f_{X_h}(x)=f_{X_0}(x)+f_{X_0}(x)\sum_{r=1}^{R}\frac{h^r}{r!}P_r(x)+\mathcal{O}(h^{R+\frac{1}{2}}) \tag{24}$$

uniformly with respect to $x\in\mathbf{R}$.
(*b*) *If furthermore* $\sup_{h\in\mathcal{H},x,v\in\mathbf{R}}\left|\partial_x^{(2R)}f_{X_0|\Delta_h=v}(x)\right|<+\infty$ *and* $\lim_{x\to-\infty}\partial_x^{(2R)}f_{X_0|\Delta_h=v}(x)=0$ *for every* $v\in\mathbf{R}$, *then*

$$F_{X_h}(x)=F_{X_0}(x)+\sum_{r=1}^{R}\frac{h^r}{r!}\mathbf{E}\left[P_r(X_0)\mathbf{1}_{\{X_0\leqslant x\}}\right]+\mathcal{O}(h^{R+\frac{1}{2}}) \tag{25}$$

uniformly with respect to $x\in\mathbf{R}$.

Proof. The case $R=0$ is trivial, using the expansion (6) and the convention $\sum_{r=1}^{0}=0$. Let $g:\mathbf{R}\to\mathbf{R}$ be an infinitely differentiable test function with compact support. We apply the expansion (6) to the smooth function g where coefficients $c(r,2r+1)$, $r=1,\dots,R-1$ and $c(r,2R)$, $r=R,\dots,2R-1$ are given by (7) and the remainder term $\mathcal{R}^g_{2R+1}$ is given by (8).

We first note that, for every $\ell\in\{1,\dots,2R\}$,

$$\mathbf{E}\left[g^{(\ell)}(X_0)b_{r,\ell-r}(Y)\right]=\int_{\mathbf{R}^2}g^{(\ell)}(x)b_{r,\ell-r}(y)f_{X_0,Y}(x,y)\mathrm{d}x\mathrm{d}y.$$

Then, performing successively ℓ integrations by parts yields

$$\mathbf{E}\left[g^{(\ell)}(X_0)b_{r,\ell-r}(Y)\right]=\int_{\mathbf{R}}g(x)\int_{\mathbf{R}}(-1)^{\ell}b_{r,\ell-r}(y)\partial_x^{(\ell)}f_{X_0,Y}(x,y)\mathrm{d}y\mathrm{d}x.$$

As for the remainder term,

$$\begin{aligned}\mathcal{R}^g_{2R+1}&=\frac{1}{(2R)!}\mathbf{E}\left[\int_0^{X_h-X_0}g^{(2R+1)}(t+X_0)(X_h-X_0-t)^{2R}\mathrm{d}t\right]\\&=\frac{1}{(2R)!}\mathbf{E}\left[\int_0^1 g^{(2R+1)}(X_0+s\Delta_h)(\Delta_h)^{2R+1}(1-s)^{2R}\mathrm{d}s\right]\\&=\frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}^2}g^{(2R+1)}(x)v^{2R+1}f_{X_0,\Delta_h}(x-sv,v)\mathrm{d}x\mathrm{d}v(1-s)^{2R}\mathrm{d}s.\end{aligned}$$

Performing successively $2R+1$ integrations by parts yields

$$\begin{aligned}\mathcal{R}^g_{2R+1}&=\frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}}\left(\int_{\mathbf{R}}g(x)\partial_x^{(2R+1)}f_{X_0,\Delta_h}(x-sv,v)\mathrm{d}x\right)v^{2R+1}\mathrm{d}v(1-s)^{2R}\mathrm{d}s,\\&=\int_{\mathbf{R}}g(x)r(h,x)\mathrm{d}x,\end{aligned}$$

where

$$\begin{aligned}r(h,x)&=\frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}}\partial_x^{(2R+1)}f_{X_0,\Delta_h}(x-sv,v)v^{2R+1}\mathrm{d}v(1-s)^{2R}\mathrm{d}s,\\&=\frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}}\partial_x^{(2R+1)}f_{X_0|\Delta_h=v}(x-sv)f_{\Delta_h}(v)v^{2R+1}\mathrm{d}v(1-s)^{2R}\mathrm{d}s.\end{aligned} \tag{26}$$

Plugging these identities in (6), we get that, for every test-function g,

$$\mathbf{E}\left[g(X_h)\right] = \int_{\mathbf{R}^d} g(x)\left[f_{X_0}(x) + f_{X_0}(x)\sum_{r=1}^{R}\frac{h^r}{r!}P_r(x) + \widetilde{r}(h,x)\right]\mathrm{d}x,$$

where

$$\widetilde{r}(h,x) = f_{X_0}(x)\sum_{r=R+1}^{2R-1}\frac{h^r}{r!}P_r(x) + r(h,x),$$

Hence

$$f_{X_h}(x) = f_{X_0}(x) + f_{X_0}(x)\sum_{r=1}^{R}\frac{h^r}{r!}P_r(x) + \widetilde{r}(h,x). \tag{27}$$

The continuity of the function on the right hand side of the above equality will establish the announced expansion, provided we show that $r(h,x) = \mathcal{O}\big(h^{R+\frac{1}{2}}\big)$ uniformly with respect to $x \in \mathbf{R}$. It follows from the boundedness assumption made on $\partial_x^{(2R+1)} f_{X_0|\Delta_h=v}(x)$ that

$$\big|r(h,x)\big| \leqslant \frac{1}{(2R)!}\sup_{x,v\in\mathbf{R}}\left|\partial_x^{(2R+1)} f_{X_0|\Delta_h=v}(x)\right|\mathbf{E}\left[\frac{\left|\Delta_h\right|^{2R+1}}{2R+1}\right] \leqslant \frac{C_{\Xi,R}}{(2R+1)!}h^{R+\frac{1}{2}}, \tag{28}$$

owing to the upper-bound established in (17) for $\mathbf{E}\left[|X_h - X_0|^{2R+1}\right]$, since $\Xi \in L^{2R+1}$.
(b) The claim amounts to integrating Equation (27), provided we show that the integrals of $P_r(x)f_{X_0}(x)$, for $r = 1,\ldots,2R-1$, are at least semi-convergent and that, for all $b \in \mathbf{R}$, $\int_{-\infty}^{b} r(h,x)\mathrm{d}x = \mathcal{O}(h^{R+\frac{1}{2}})$. Owing to Fubini's Theorem, using the definition (26) of $r(h,x)$, we have for all $a < b \in \mathbf{R}$,

$$\int_a^b r(h,x)\mathrm{d}x = \frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}}\left(\partial_x^{(2R)} f_{X_0|\Delta_h=v}(b-sv) - \partial_x^{(2R)} f_{X_0|\Delta_h=v}(a-sv)\right) \\ \times f_{\Delta_h}(v)v^{2R+1}\mathrm{d}v(1-s)^{2R}\mathrm{d}s$$

The assumption $\sup_{h\in\mathcal{H},x,v\in\mathbf{R}}|f^{(2R)}_{X_0|\Delta_h=v}(x)| < +\infty$ and the upper bound (17), yield that $\int_a^b \big|r(h,x)\big|\mathrm{d}x < +\infty$. Hence, owing to Lebesgue's Dominated Convergence Theorem and to the assumption
$\lim_{x\to-\infty} f^{(2R)}_{X_0|\Delta_h=v}(x) = 0$, we get

$$\begin{aligned}\int_{-\infty}^b r(h,x)\mathrm{d}x &= \frac{1}{(2R)!}\int_0^1\int_{\mathbf{R}}\partial_x^{(2R)} f_{X_0|\Delta_h=v}(b-sv)f_{\Delta_h}(v)v^{2R+1}\mathrm{d}v(1-s)^{2R}\mathrm{d}s,\\ &= \frac{1}{(2R)!}\int_0^1\mathbf{E}\left[\partial_x^{(2R)} f_{X_0|\Delta_h}(b-s\Delta_h)(\Delta_h)^{2R+1}\right](1-s)^{2R}\mathrm{d}s.\end{aligned} \tag{29}$$

and then, likewise (28), using the boundedness of $f^{(2R)}_{X_0\,|\,\Delta_h=v}(x)$,

$$\int_{-\infty}^b r(h,x)\mathrm{d}x = \mathcal{O}(h^{R+\frac{1}{2}}). \tag{30}$$

Owing to Equation (27), if we take $h_1, \ldots, h_{2R-1} \in \mathcal{H}$ pairwise distinct, we get, for all $i = 1, \ldots, 2R-1$,

$$\sum_{r=1}^{2R-1} h_i^{r-1} \left(\frac{P_r(x)}{r!} f_{X_0}(x) \right) = \rho_i(x), \tag{31}$$

with

$$\rho_i(x) = \frac{f_{X_{h_i}}(x) - f_{X_0}(x) - r(h_i, x)}{h_i}, \quad i = 1, \ldots, 2R-1. \tag{32}$$

Hence we get a Vandermonde system, $Vu(x) = \rho(x)$ with

$$V = V(h_1, \ldots, h_{2R-1}) = \begin{pmatrix} 1 & h_1 & h_1^2 & \ldots & h_1^{2R-2} \\ 1 & h_2 & h_2^2 & \ldots & h_2^{2R-2} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & h_{2R-1} & h_{2R-1}^2 & \ldots & h_{2R-1}^{2R-2} \end{pmatrix},$$

$$u(x) = (u_1(x), \ldots, u_{2R-1}(x)) \quad \text{with} \quad u_r(x) = \frac{P_r(x)}{r!} f_{X_0}(x)$$

and

$$\rho(x) = (\rho_1(x), \ldots, \rho_{2R-1}(x)).$$

We set, for all $j = 1, \ldots, 2R-1$,

$$\widetilde{V}_j(x) = \widetilde{V}_j(h_1, \ldots, h_{2R-1}, \rho(x)) = \begin{pmatrix} 1 & \ldots & h_1^{j-2} & \rho_1(x) & h_1^j & \ldots & h_1^{2R-2} \\ 1 & \ldots & h_2^{j-2} & \rho_2(x) & h_2^j & \ldots & h_2^{2R-2} \\ \vdots & \ldots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & \ldots & h_{2R-1}^{j-2} & \rho_{2R-1}(x) & h_{2R-1}^j & \ldots & h_{2R-1}^{2R-2} \end{pmatrix}.$$

By expanding along the j^{th} column, the determinant of $\widetilde{V}_j$ writes

$$\det(\widetilde{V}_j(x)) = \sum_{i=1}^{2R-1} (-1)^{i+j} d_{ij} \rho_i(x),$$

where $d_{ij} := d_{ij}(h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_{2R}) \in \mathbf{R}$ with

$$d_{ij}(h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_{2R-1}) = \det \begin{pmatrix} 1 & \ldots & h_1^{j-2} & h_1^j & \ldots & h_1^{2R-2} \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots \\ 1 & \ldots & h_{i-1}^{j-2} & h_{i-1}^j & \ldots & h_{i-1}^{2R-2} \\ 1 & \ldots & h_{i+1}^{j-2} & h_{i+1}^j & \ldots & h_{i+1}^{2R-2} \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots \\ 1 & \ldots & h_{2R-1}^{j-2} & h_{2R-1}^j & \ldots & h_{2R-1}^{2R-2} \end{pmatrix}.$$

Hence, owing to Cramer's rule, the solution of the Vandermonde system writes

$$u_j(x) = \frac{\det\left(\widetilde{V}_j(x)\right)}{\det(V)} = \frac{1}{\det(V)} \sum_{i=1}^{2R-1} (-1)^{i+j} d_{ij} \rho_i(x).$$

Finally, since we saw that for all $i = 1, \ldots, 2R-1$, the integral of $\rho_i(x)$ is semi-convergent, we deduce the semi-convergence of the integral

$$\int_{-\infty}^{b} P_r(x) f_{X_0}(x)\mathrm{d}x = \frac{r!}{\det(V)} \sum_{i=1}^{2R-1} (-1)^{i+r} d_{ir} \int_{-\infty}^{b} \rho_i(x)\mathrm{d}x,$$

where, owing to the expression (32) and to (30), $\int_{-\infty}^{b} \rho_i(x)\mathrm{d}x$ is finite, which concludes the proof.

□

5.2 Strong convergence rate

We conclude by showing the strong convergence rate (SE_β) for the nested Monte Carlo estimator. The following Lemma is more or less standard (see for instance [Avi09]).

Lemma 5.1. *Let ξ and ξ' be two real valued random variables lying in L^p, $p \geqslant 1$, with densities f_ξ and $f_{\xi'}$ respectively. Then, for every $x \in \mathbf{R}$,*

$$\left\|\mathbf{1}_{\{\xi \leqslant x\}} - \mathbf{1}_{\{\xi' \leqslant x\}}\right\|_2^2 \leqslant \left(p^{\frac{p}{p+1}} + p^{\frac{1}{p+1}}\right) \left(\left\|f_\xi\right\|_{\sup} + \left\|f_{\xi'}\right\|_{\sup}\right)^{\frac{p}{p+1}} \left\|\xi - \xi'\right\|_p^{\frac{p}{p+1}}. \tag{33}$$

Proof. Let $L > 0$. Note that

$$\begin{aligned}
\left\|\mathbf{1}_{\{\xi \leqslant x\}} - \mathbf{1}_{\{\xi' \leqslant x\}}\right\|_2^2 &= \mathbf{P}\left(\xi \leqslant x \leqslant \xi'\right) + \mathbf{P}\left(\xi' \leqslant x \leqslant \xi\right) \\
&\leqslant \mathbf{P}\left(\xi \leqslant x,\, \xi' \geqslant x + L\right) + \mathbf{P}\left(\xi \leqslant x \leqslant \xi' \leqslant x + L\right) \\
&\quad + \mathbf{P}\left(\xi' \leqslant x, \xi \geqslant x + L\right) + \mathbf{P}\left(\xi' \leqslant x \leqslant \xi \leqslant x + L\right) \\
&\leqslant \mathbf{P}\left(\xi' - \xi \geqslant L\right) + \mathbf{P}\left(\xi - \xi' \geqslant L\right) \\
&\quad + \mathbf{P}(\xi' \in [x, x+L]) + \mathbf{P}(\xi \in [x, x+L]) \\
&= \mathbf{P}\left(|\xi' - \xi| \geqslant L\right) + \mathbf{P}(\xi \in [x, x+L]) + \mathbf{P}(\xi' \in [x, x+L]) \\
&\leqslant \frac{\mathbf{E}\left[|\xi' - \xi|^p\right]}{L^p} + L\Big(\left\|f_\xi\right\|_{\sup} + \left\|f_{\xi'}\right\|_{\sup}\Big).
\end{aligned}$$

A straightforward optimization in L yields the announced result. □

The strong convergence is a consequence of the Proposition 5.1 combined with the previous Lemma and Lemma 3.2.

Proposition 5.2. *Assume $X \in L^p$, $p \geqslant 2$. Under the assumptions of Proposition 5.1 (a) with $R = 0$ and if the density f_{X_0} is bounded, then there exists $h_0 = \frac{1}{K_0} \in \mathcal{H} \setminus \{0\}$ such that, for every $h, h' \in (0, h_0]$*

$$\left\|\mathbf{1}_{\{X_h \leqslant x\}} - \mathbf{1}_{\{X_{h'} \leqslant x\}}\right\|_2^2 \leqslant C|h - h'|^{\frac{p}{2(p+1)}},$$

where $C = 2^{\frac{p}{p+1}}\left(p^{\frac{p}{p+1}} + p^{\frac{1}{p+1}}\right)(\left\|f_{X_0}\right\|_{\sup} + 1)^{\frac{p}{p+1}}\left(4\, B_p \left\|X\right\|_p\right)^{\frac{p}{p+1}}$. This means that the strong approximation error assumption holds with $\beta = \frac{p}{2(p+1)} \in (0, \frac{1}{2})$.

Proof. It follows from Proposition 5.1 (a) that

$$f_{X_h}(x) \leqslant f_{X_0}(x) + o(h^{\frac{1}{2}}) \quad \text{uniformly with respect to } x \in \mathbf{R}.$$

Consequently, there exists an $h_0 = \frac{1}{K_0} \in \mathcal{H} \setminus \{0\}$ such that, for every $h \in (0, h_0]$,

$$\forall\, x \in \mathbf{R},\ f_{X_h}(x) \leqslant f_{X_0}(x) + 1.$$

Plugging the above bound and (14) in Inequality (33) of Lemma 5.1 applied with $\xi = X_h$ and $\xi' = X_{h'}$ completes the proof. $\square$

References

[Avi09] R. Avikainen. "On irregular functionals of SDEs and the Euler scheme". In: *Finance and Stochastics* 13.3 (2009), pp. 381–401.

[BHR15] K. Bujok, B.M. Hambly, and C. Reisinger. "Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives". In: *Methodology and Computing in Applied Probability* 17.3 (2015), pp. 579–604.

[BT96a] V. Bally and D. Talay. "The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function". In: *Probab. Theory Related Fields* 104.1 (1996), pp. 43–60.

[BT96b] V. Bally and D. Talay. "The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density". In: *Monte Carlo Methods Appl.* 2.2 (1996), pp. 93–128.

[CL12] N. Chen and Y. Liu. "Estimating expectations of functionals of conditional expected via multilevel nested simulation". In: *Presentation at conference on Monte Carlo and Quasi-Monte Carlo Methods, Sydney.* 2012.

[Com74] L. Comtet. *Advanced combinatorics.* enlarged. The art of finite and infinite expansions. D. Reidel Publishing Co., Dordrecht, 1974, pp. xi+343.

[DL09] L. Devineau and S. Loisel. "Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique Solvabilité II". In: *Bulletin Français d'Actuariat, Institut des Actuaires, 10 (17)* (2009), pp. 188–221.

[Gil08] M.B. Giles. "Multilevel Monte Carlo path simulation". In: *Oper. Res.* 56.3 (2008), pp. 607–617.

[Gil15] M.B. Giles. "Multilevel Monte Carlo methods". In: *Acta Numer.* 24 (2015), pp. 259–328.

[Gio17] D. Giorgi. "Théorèmes limites pour estimateurs Multilevel avec et sans poids. Comparaisons et applications." PhD thesis. Université Pierre et Marie Curie, 2017.

[GJ10] M.B. Gordy and S. Juneja. "Nested Simulation in Portfolio Risk Measurement". In: *Management Science 56(10):1833-1848* (2010).

[GLP17] D. Giorgi, V. Lemaire, and G. Pagès. "Limit theorems for weighted and regular multilevel estimators". In: *Monte Carlo Methods Appl.* 23.1 (2017), pp. 43–70.

[Haj12] Abdul L. Haji A. "Pedestrian Flow in the Mean Field Limit". MA thesis. King Abdullah University of Science and Technology, Thuwal, 2012.

[LP17] V. Lemaire and G. Pagès. "Multilevel Richardson–Romberg extrapolation". In: *Bernoulli* 23.4A (2017), pp. 2643–2692.

[Pag18] G. Pagès. *Numerical Probability: An Introduction with Applications to Finance.* Universitext. Springer, 2018, p. 607.

[Shi96] A. N. Shiryaev. *Probability.* Second. Vol. 95. Graduate Texts in Mathematics. Translated from the first (1980) Russian edition by R. P. Boas. New York: Springer-Verlag, 1996, pp. xvi+623.

[TT90] D. Talay and L. Tubaro. "Expansion of the global error for numerical schemes solving stochastic differential equations". In: *Stochastic Anal. Appl.* 8.4 (1990), 483–509 (1991).