Central limit theorems for multilevel Monte Carlo methods

Håkon Hoel[a,*], Sebastian Krumscheid[b]

[a] *Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden*
[b] *Calcul Scientifique et Quantification de l'Incertitude (CSQI), Institute of Mathematics, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

Abstract

In this work, we show that uniform integrability is not a necessary condition for central limit theorems (CLT) to hold for normalized multilevel Monte Carlo (MLMC) estimators and we provide near optimal weaker conditions under which the CLT is achieved. In particular, if the variance decay rate dominates the computational cost rate (i.e., $\beta > \gamma$), we prove that the CLT applies to the standard (variance minimizing) MLMC estimator. For other settings where the CLT may not apply to the standard MLMC estimator, we propose an alternative estimator, called the mass-shifted MLMC estimator, to which the CLT always applies. This comes at a small efficiency loss: the computational cost of achieving mean square approximation error $\mathcal{O}(\epsilon^2)$ is at worst a factor $\mathcal{O}(\log(1/\epsilon))$ higher with the mass-shifted estimator than with the standard one.

Keywords: Multilevel Monte Carlo, Central Limit Theorem

1. Introduction

The multilevel Monte Carlo (MLMC) method is a hierarchical sampling method which in many settings improves the computational efficiency of weak approximations by orders of magnitude. The method was independently introduced in the papers [19, 13] for the purpose of parametric integration and for approximations of observables of stochastic differential

*Corresponding author
Email addresses: hhakon@chalmers.se, haakonah1@gmail.com (Håkon Hoel), sebastian.krumscheid@epfl.ch (Sebastian Krumscheid)

equations, respectively. MLMC methods have since been applied with considerable success in a vast range of stochastic problems, a collection of which can be found in the overview [14]. In this work we present near optimal conditions under which the normalized MLMC estimator converges in distribution to a standard normal distribution. Our result has applications in settings where the MLMC approximation error is measured in terms of probability of failure (6) rather than the classical mean square error.

1.1. Main result

We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $X \in L^2(\Omega)$ be a scalar random variable (r. v.) for which we seek the expectation $\mathbb{E}[X]$. Let $\{X_\ell\}_{\ell=-1}^{\infty} \subset L^2(\Omega)$ be a sequence of r. v. satisfying the following:

Assumption 1.1. There exist rate constants $\alpha, \beta, \gamma > 0$ with $\min(\beta, \gamma) \le 2\alpha$ and a constant $c_\alpha > 0$ such that

(i) $|\mathbb{E}[X - X_\ell]| \le c_\alpha 2^{-\alpha\ell}$ for all $\ell \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$,

(ii) $V_0 > 0$ and $V_\ell := \mathrm{Var}(\Delta_\ell X) = \mathcal{O}_\ell(2^{-\beta\ell})$,

(iii) $C_\ell := \mathrm{Cost}(\Delta_\ell X) = \Theta_\ell\left(2^{\gamma\ell}\right)$,

where $\Delta_\ell X := X_\ell - X_{\ell-1}$ with $X_{-1} := 0$. The notation $f(x_\ell) = \mathcal{O}_\ell(y_\ell)$ means there exists a constant $C > 0$ such that $|f(x_\ell)| < C|y_\ell|$ for all $\ell \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $f(x_\ell) = \Theta_\ell(y_\ell)$ means there exist constants $C > c > 0$ such that $c|y_\ell| < |f(x_\ell)| < C|y_\ell|$ for all $\ell \in \mathbb{N}_0$.

Definition 1.1 (Variance minimizing MLMC estimator [14, 19])**.** The MLMC estimator $\mathcal{A}_{ML} \colon (0, \infty) \to L^2(\Omega)$ applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumption 1.1 is defined by

$$\mathcal{A}_{ML}(\epsilon) = \sum_{\ell=0}^{L(\epsilon)} \sum_{i=1}^{M_\ell(\epsilon)} \frac{\Delta_\ell X^i}{M_\ell(\epsilon)} .$$

Here

$$L^2(\Omega) \ni \Delta_\ell X^i = X_\ell^i - X_{\ell-1}^i, \quad \ell \in \mathbb{N}_0, \quad i \in \mathbb{N}$$

denotes a sequence of independent r.v. and every subsequence $\{\Delta_\ell X^i\}_i$ consist of independent and identically distributed (i.i.d.) r.v., the number of levels is

$$L(\epsilon) := \max\left(\left\lceil \frac{\log_2(c_\alpha \epsilon^{-1})}{\alpha} \right\rceil, 1\right), \quad \epsilon > 0, \tag{1}$$

and the number of samples per level $\ell = 0, 1, \dots$ is

$$M_\ell(\epsilon) := \max\left(\left\lceil \epsilon^{-2}\sqrt{\frac{V_\ell}{C_\ell}} S_{L(\epsilon)} \right\rceil, 1\right), \quad \epsilon > 0\ , \tag{2}$$

with the monotonically increasing sequence S_k defined as

$$S_k := \sum_{\ell=0}^{k} \sqrt{V_\ell C_\ell}, \quad k \in \mathbb{N}_0\ . \tag{3}$$

For any fixed and sufficiently large computational budget $c > 0$, the sequence $\{M_\ell\}_{\ell=0}^{L}$ in (2) is the one in $\mathbb{N}^L$ that minimizes $\mathrm{Var}(\mathcal{A}_{ML})$ subject the constraint $\mathrm{Cost}(\mathcal{A}_{ML}) \leq c$, cf. [13]. We will therefore refer to $\mathcal{A}_{ML}$ as the variance minimizing MLMC estimator.

It is known that MLMC estimators can offer significant complexity (i.e., cost vs. accuracy) benefits compared to classic Monte Carlo estimators [14]. In fact, the variance minimizing estimator $\mathcal{A}_{ML}(\epsilon)$ reduces the computational cost for achieving an approximation with mean square error of $\mathcal{O}_\epsilon\big(\epsilon^2\big)$ from $\Theta_\epsilon\big(\epsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}\big)$ for the classic Monte Carlo method to $\Theta_\epsilon\big(\epsilon^{-2} S_{L(\epsilon)}^2 + C_{L(\epsilon)}\big)$, where

$$S_{L(\epsilon)} = \begin{cases} \mathcal{O}_\epsilon(1) & \text{if} \quad \beta > \gamma\ , \\ \mathcal{O}_\epsilon\big(\log\big(\epsilon^{-1}\big)\big) & \text{if} \quad \beta = \gamma\ , \\ \mathcal{O}_\epsilon\big(\epsilon^{-\frac{\gamma-\beta}{2\alpha}}\big) & \text{if} \quad \beta < \gamma\ , \end{cases}$$

and $C_{L(\epsilon)} = \Theta_\epsilon(\epsilon^{-\gamma/\alpha})$ as functions of the rate triplet introduced in Assumption 1.1.

In this work, we address the asymptotic normality of the MLMC estimator. For convenience, we will refer to

$$\frac{\mathcal{A}_{ML}(\epsilon) - \mathbb{E}\big[X_{L(\epsilon)}\big]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML}(\epsilon))}}$$

as the normalized estimator. When confusion is not possible, we will use the following shorthands,

$$\mathcal{A}_{ML} := \mathcal{A}_{ML}(\epsilon)\ , \quad M_\ell := M_\ell(\epsilon)\ , \quad L := L(\epsilon).$$

The following conventions will be employed throughout this work:

$$0 \cdot (\pm\infty) = 0 \quad \text{and} \quad 0/0 = 0\ .$$

We are ready to state the main result of this work.

Theorem 1.1 (Main result)**.** *Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumption 1.1. Additionally, if*

(i) $\beta > \gamma$, impose no further assumptions,

(ii) $\gamma \geq \beta$ and $\lim_{\ell\to\infty} S_\ell < \infty$, impose no further assumptions,

(iii) $\beta = \gamma$ and $\lim_{\ell\to\infty} S_\ell = \infty$, assume that

$$\lim_{\ell\to\infty} \mathbf{1}_{\{V_\ell>0\}} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \nu S_\ell^2 \exp((2\alpha-\gamma)\ell)\right\}} \right] = 0 \quad \forall \nu > 0, \tag{4}$$

(iv) $\gamma > \beta$ and $\lim_{\ell\to\infty} S_\ell = \infty$, assume that $\beta < 2\alpha$, equality (4) *holds and that there exists an $\upsilon \in [\beta, 2\alpha)$ such that $\lim_{k\to\infty} S_k 2^{(\upsilon-\gamma)k/2} > 1$.*

Then the normalized estimator satisfies the central limit theorem (CLT), in the sense that

$$\frac{\mathcal{A}_{ML} - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \quad \textit{as} \quad \epsilon \downarrow 0. \tag{5}$$

The main result follows from Theorems 2.5 and 2.6. We note that Theorem 1.1 in particular implies that the CLT always applies to the normalized variance minimizing MLMC estimator when $\beta > \gamma$.

Remark 1.1. The reason why we have not included the setting $\gamma > \beta$ and $\beta = 2\alpha$ in Theorem 1.1 is that one cannot impose reasonable assumptions to exclude $M_L = \Theta_\epsilon(1)$ and $V_L/\mathrm{Var}(\mathcal{A}_{ML}) = \Theta_\epsilon(1)$; cf. Example 2.1. In such cases, a non-negligible contribution to the variance of the normalized estimator may derive from a finite number of samples on the finer levels $L, L-1, \ldots$. For example, if $M_L = 1$ and $V_L/\mathrm{Var}(\mathcal{A}_{ML}) \geq c > 0$ for all $\epsilon > 0$ sufficiently small, then

$$\frac{\mathcal{A}_{ML} - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} = \sum_{\ell=0}^{L-1} \left(\sum_{i=1}^{M_\ell} \frac{\Delta_\ell X^i}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})} M_\ell} \right) + \frac{\Delta_L X^1 - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} ,$$

and the CLT applies only if $\Delta_\ell X$ converges in distribution to a Gaussian as $\ell \to \infty$.

1.2. Probability of failure

Distributional properties of normalized sample estimators can be useful for controlling the probability of (approximation) failure:

$$\mathbb{P}(|\mathcal{A}_{ML} - \mathbb{E}[X]| \geq 2\epsilon) \leq \delta \, . \tag{6}$$

Here, $2\epsilon > 0$ denotes the accuracy and $1-\delta > 0$ the confidence. To control the probability of failure, one may dominate the total error from above by the sum of a bias and a statistical error:

$$\mathbb{P}(|\mathcal{A}_{ML} - \mathbb{E}[X]| \geq 2\epsilon) \leq \mathbb{P}(|\mathbb{E}[X_L] - \mathbb{E}[X]| \geq \epsilon) + \mathbb{P}(|\mathcal{A}_{ML} - \mathbb{E}[X_L]| \geq \epsilon) \ . \tag{7}$$

Assumption 1.1(i) and the value of L ensure that the bias constraint is met

$$|\mathbb{E}[X_L] - \mathbb{E}[X]| \leq \epsilon.$$

Supposing next that the CLT applies, the key step in (approximately) controlling the statistical error is the approximation

$$\frac{\mathcal{A}_{ML} - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} \overset{\mathrm{d}}{\approx} \mathcal{N}(0,1).$$

The use of CLT in efficient algorithms for controlling the probability of failure is a motivation for the goal of this work: to describe as weak as possible conditions under which the CLT applies to the standard MLMC estimator.

Remark 1.2. Whenever $\beta \geq \gamma$ and $\alpha > \gamma/2$, one may reduce the bias of the variance minimizing MLMC estimator without affecting the asymptotic growth rate of the computational cost by replacing the rate parameter α by $\gamma/2$ in the formula for L in (1) and updating the values for $\{M_\ell\}_{\ell=0}^L$ accordingly. This replacement leads to an asymptotically vanishing bias to standard deviation ratio,

$$\lim_{\epsilon \downarrow 0} \frac{\mathbb{E}[X_L] - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} = \lim_{\epsilon \downarrow 0} \epsilon^{2\alpha/\gamma - 1} = 0,$$

and it relates to an uneven splitting of the accuracy between the bias and the statistical error constraints in (7). That is,

$$\begin{aligned}\mathbb{P}(|\mathcal{A}_{ML} - \mathbb{E}[X]| \geq 2\epsilon) \leq \mathbb{P}(&|\mathbb{E}[X_L] - \mathbb{E}[X]| \geq \theta(\epsilon)\epsilon) \\ &+ \mathbb{P}(|\mathcal{A}_{ML} - \mathbb{E}[X_L]| \geq (2 - \theta(\epsilon))\epsilon)\end{aligned}$$

for any monotonically increasing function $\theta : (0, \infty) \to (0, 1]$ satisfying $\theta(\epsilon) \geq (\epsilon/c_\alpha)^{2\alpha/\gamma - 1}$, cf. [9]. We leave as a remark that by straightforward extension of Theorem 1.1, the CLT also applies to the normalized variance minimizing MLMC estimator with θ-splitting in settings where $\beta \geq \gamma$ and Theorem 1.1's assumptions hold.

1.3. The mass-shifted MLMC estimator

In [28, 31, 32] Glynn et al. show that for a collection of r.v. $\{X_\ell\}_{\ell=-1}^{\infty}$ satisfying Assumption 1.1 one can construct the following unbiased coupled sampling method for the limit r.v. X:

$$Z = \sum_{\ell=0}^{\infty} \frac{\Delta_\ell X \; \mathbf{1}_{\{N \geq \ell\}}}{\mathbb{P}(N \geq \ell)}.$$

Here, the r.v. $N : \Omega \to \mathbb{N}_0$ is independent of $\{\Delta_\ell X\}_{\ell=-1}^{\infty}$ and $\mathbb{P}(N \geq \ell) > 0$ for all $\ell \geq 0$. Provided N is chosen such that $\mathbb{E}[|Z|] < \infty$, the strong law of large numbers yields that

$$\overline{Z}_M = \frac{1}{M} \sum_{i=0}^{M} Z^i \overset{\text{a.s.}}{\to} \mathbb{E}[X] \quad \text{as} \quad M \to \infty,$$

where $Z^1, Z^2, \ldots$ is an i.i.d. sequence with $Z^i \overset{d}{=} Z$. Although $\overline{Z}_M$ clearly is not an MLMC estimator of the kind studied in this paper, one may view it, when the number of samples M is large, as a randomized MLMC estimator where both L and $M_\ell \approx M \times \mathbb{P}(N \geq \ell)$ for all $\ell \geq 0$ are random non-negative numbers, cf. [31]. By carefully choosing the distribution of N such that $\mathrm{Var}(Z^i) < \infty$ and exploiting that $\overline{Z}_M$ is the sum of i.i.d. random variables, Glynn et al. prove that the CLT applies to $(\overline{Z}_M - \mathbb{E}[X])/\sqrt{\mathrm{Var}(\overline{Z}_M)}$ in settings where $\beta \geq \gamma$.

Concerning the efficiency of the method, it can be shown that the distribution N that minimizes the quantity $\mathrm{Var}(\overline{Z}_M) \times \mathrm{Cost}(\overline{Z}_M)$, satisfies

$$\mathbb{P}(N \geq \ell) = \Theta_\ell(\sqrt{V_\ell / C_\ell}) \tag{8}$$

(supposing, unlike our approach, that $V_\ell > 0$ for all ℓ). When $\beta > \gamma$, any distribution N satisfying (8) induces a distribution Z that has bounded variance, and consequently, the CLT applies. When $\beta = \gamma$, however, it turns out that $\mathrm{Var}(Z) = \infty$ for any N satisfying (8), so that in order to obtain the CLT one needs to consider distributions N whose mass is shifted slightly from the efficiency optimizing (8) to the tail:

$$\mathbb{P}(N \geq \ell) = \Theta_\ell((\ell+1)\log(\ell+2)^{1+\xi}\sqrt{V_\ell/C_\ell}), \qquad \xi > 0.$$

This shift leads to an estimator $\overline{Z}_M$ with approximation error $\mathbb{E}\left[\left(\overline{Z}_M - \mathbb{E}[Z]\right)^2\right] = \mathcal{O}_\epsilon(\epsilon^2)$ obtained at the (random) computational cost $\mathcal{O}_\epsilon(\epsilon^{-2}\log(1/\epsilon)^2 \log(\log(1/\epsilon))^{1+\xi})$.

In comparison, for the settings covered by Theorem 1.1 when $\beta = \gamma$, the variance minimizing estimator $\mathcal{A}_{ML}(\epsilon)$ achieves the MSE $\mathcal{O}_\epsilon(\epsilon^2)$ at the slightly lower (and non-random) computational cost $\Theta_\epsilon(\epsilon^{-2} S_L^2) = \mathcal{O}_\epsilon(\epsilon^{-2}\log(1/\epsilon)^2)$.

Taking inspiration of from Glynn et al.'s mass-shifting approach, we propose the following relative shift of "sample mass" from the lower levels of the variance minimizing estimator's optimal $\{M_\ell\}_{\ell=0}^L$ to the higher levels:

$$\widetilde{M}_\ell := \max\left(\left\lceil \epsilon^{-2}(S_\ell+1)\log(S_\ell+1)^{1+\xi}\sqrt{\frac{V_\ell}{C_\ell}}\widetilde{S}_L \right\rceil, 1\right), \tag{9}$$

where

$$\widetilde{S}_L := \sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{(S_\ell+1)\log(S_\ell+1)^{1+\xi}}, \quad \xi > 0,$$

and the resulting estimator

$$\widetilde{\mathcal{A}}_{ML} = \sum_{\ell=0}^{L}\sum_{i=1}^{\widetilde{M}_\ell} \frac{\Delta_\ell X^i}{\widetilde{M}_\ell}. \tag{10}$$

We will refer to $\widetilde{\mathcal{A}}_{ML}$ as the mass-shifted MLMC estimator. The CLT applies in all relevant settings for the normalized version of this estimator:

Theorem 1.2 (CLT for mass-shifted MLMC)**.** *For any $\xi > 0$, let $\widetilde{\mathcal{A}}_{ML}$ denote the resulting mass-shifted MLMC estimator applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumption 1.1. Then the normalized mass-shifted MLMC estimator satisfies*

$$\frac{\widetilde{\mathcal{A}}_{ML} - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}\left(\widetilde{\mathcal{A}}_{ML}\right)}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \quad \text{as} \quad \epsilon \downarrow 0 \tag{11}$$

and the approximation error $\mathbb{E}\left[\left(\widetilde{\mathcal{A}}_{ML} - \mathbb{E}[X]\right)^2\right] = \mathcal{O}(\epsilon^2)$ is achieved at the computational cost

$$\Theta_\epsilon(\epsilon^{-2}(S_L+1)^2\log(S_L+1)^{1+\xi}) = \begin{cases} \mathcal{O}_\epsilon(\epsilon^{-2}), & \beta > \gamma \\ \mathcal{O}_\epsilon(\epsilon^{-2}\log(1/\epsilon)^2\log(\log(1/\epsilon))^{1+\xi}), & \beta = \gamma \\ \mathcal{O}_\epsilon(\epsilon^{-2-\frac{\gamma-\beta}{\alpha}}\log(1/\epsilon)^{1+\xi}), & \gamma > \beta. \end{cases}$$

The proof of Theorem 1.2 is given in Section 2.1.

1.4. Literature review

In addition to the above mentioned contributions by Glynn et al., the CLT has been proved for MLMC methods through assuming (or verifying for the particular sequence of r.v. considered) either a Lyapunov condition [18], or uniform integrability [1, 10, 15], or a weaker higher moment decay rate [9] for the sequence $\{\mathbf{1}_{\{V_\ell>0\}}|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2/V_\ell\}_{\ell\in\mathbb{N}_0}$. To show that this work extends the existing literature, we now provide an explicit example that is covered by Theorem 1.1 but where uniform integrability does not hold.

Example 1.1. Consider the stochastic differential equation

$$dY = a(Y)\,dt + b(Y)\,dW(t) \qquad t\in[0,T] \tag{12}$$

with final time $T>0$, initial condition $Y(0)\in\mathbb{R}$, and coefficients $a,b:\mathbb{R}\to\mathbb{R}$ whose partial derivatives of all orders are continuous and uniformly bounded. For a given strike $K\in\mathbb{R}$, we seek to approximate the expectation of the (non-discounted) digital option payoff $X=\mathbf{1}_{\{Y(T)\geq K\}}$. Let $X_\ell = \mathbf{1}_{\{Y_\ell(T)\geq K\}}$ denote the ℓ-th resolution approximation of X where $Y_\ell(T)$ denotes the order 1.5 strong Ito-Taylor scheme [24, Ch. 10.4] numerical solution using a uniform timestep $h_\ell = 2^{-\ell}T$. In order to minimize the variance, coupled realizations $Y_\ell(\cdot,\omega)$ and $Y_{\ell-1}(\cdot,\omega)$ use the same Wiener path sampled at different resolutions. Furthermore, the scheme's fine resolution integral increments of the form

$$\Delta z_n^\ell = \int_{nh_\ell}^{(n+1)h_\ell} W(s)-W(nh_\ell)\,dt \overset{\mathrm{d}}{=} \frac{\Delta W_n^\ell h_\ell}{2} + \frac{h_\ell^{3/2}}{\sqrt{12}}\chi_n,$$

where $\chi_n\sim N(0,1)$ and $\Delta W_n^\ell = W((n+1)h_\ell)-W(nh_\ell)$ are independent, are coupled to overlapping coarse ones as follows:

$$\begin{aligned}\Delta z_n^{\ell-1} &= \Delta z_{2n}^\ell + \int_{(2n+1)h_\ell}^{2(n+1)h_\ell} W(s)-W(2nh_\ell)\,dt\\ &= \Delta z_{2n}^\ell + h_\ell \Delta W_{2n}^\ell + \int_{(2n+1)h_\ell}^{2(n+1)h_\ell} W(s)-W((2n+1)h_\ell)\,dt\\ &= \Delta z_{2n}^\ell + h_\ell\Delta W_{2n}^\ell + \Delta z_{2n+1}^\ell.\end{aligned}$$

(That is, first generate $(\Delta z_{2n}^\ell, \Delta z_{2n+1}^\ell)(\omega)$, $\Delta W_{2n}^\ell(\omega)$ and $\Delta W_{2n+1}^\ell(\omega)$, then compute the overlapping coupled coarse increment $\Delta z_n^{\ell-1}(\omega)$ by the above formula.) Assuming that the diffusion coefficient is strictly positive and $b'|_D\neq 0$ in an open domain $D\subset\mathbb{R}$ containing $Y(0)$ and K,

$$\mathbb{P}\left(|Y_\ell(T)-K|\leq h_\ell^{3/2}\right) = \mathcal{O}_\ell(h_\ell^{3/2}) \tag{13}$$

and

$$\limsup_{\ell\to\infty} \operatorname*{ess\,sup}_{\omega\in\Omega} |\Delta_\ell X(\omega) - \mathbb{E}[\Delta_\ell X]|^2 = 1. \tag{14}$$

By the order 1.5 strong order scheme, $Y_\ell(T) - Y_{\ell-1}(T) = \mathcal{O}_\ell(h_\ell^{3/2})$, which together with (13) imply that $V_\ell = \mathrm{Var}(\Delta_\ell X) = \mathcal{O}_\ell(h_\ell^{3/2})$. Lastly, since $\mathrm{Cost}(Y_\ell) = \Theta_\ell(1/h_\ell)$, the rate triplet for $\{X_\ell\}$ becomes $\alpha = 1, \beta = 3/2$ and $\gamma = 1$.

Note further that the sequence $\{\mathbf{1}_{\{V_\ell>0\}}|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2/V_\ell\}_{\ell\in\mathbb{N}_0}$ is not uniformly integrable since by (14),

$$\limsup_{\ell\to\infty} \mathbf{1}_{\{V_\ell>0\}} \frac{\operatorname{ess\,sup}_{\omega\in\Omega} |\Delta_\ell X(\omega) - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} 2^{-\beta\ell} > 0,$$

which implies that

$$\limsup_{\ell\to\infty} \mathbf{1}_{\{V_\ell>0\}} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > x\right\}}\right] = 1, \quad \text{for any } x > 0.$$

Regardless of uniform integrability, however, the CLT applies according to Theorem 1.1 in the current setting of $\beta > \gamma$.

Applications of MLMC

We conclude this section with a brief survey on the relationship between the rate parameters β and γ from Assumption 1.1 for a couple of problems which have been frequently studied.

As a first example, consider the quantity of interest (QoI) $X = \varphi(Y) \in \mathbb{R}$ with $Y : [0,T] \times \Omega \to \mathbb{R}$ denoting the solution of an SDE of the form (12). For an approximation sequence $X_\ell = \phi(Y_\ell)$, where Y_ℓ is generated by a numerical method with uniform timestep $h_\ell = 2^{-\ell}T$, one often obtains $C_\ell = \mathrm{Cost}(X_\ell) = \mathcal{O}(h_\ell^{-1})$, yielding $\gamma = 1$ (this applies for instance to the Euler–Maruyama and the Milstein schemes). The variance decay rate β is typically more sensitive, as it tends to depend on both the strong order of convergence of the numerical method and the regularity of the functional φ. If the SDE coefficients and the QoI are all sufficiently regular, then $\beta = 1$ for the Euler–Maruyama scheme and $\beta = 2$ for the Milstein scheme, but low-regularity QoIs often lead to lower-valued β. For instance, for digital and barrier options, $\beta = 1/2$ for Euler–Maruyama and $\beta = 1$ for Milstein (provided no further smoothing is applied), cf. [14, Sec. 5]. Similar reductions in the variance decay rate may occur if the SDE coefficients have low regularity or if its driving path has lower regularity than a Wiener process, cf. [6, 17].

As a second example, let the quantity of interest be $X = \varphi(u) \in \mathbb{R}$, where $u(\omega, \cdot)\colon D \to \mathbb{R}$ denotes the solution of the linear elliptic partial differential equation (PDE)

$$-\operatorname{div}\left(a(\omega, x)\nabla u(\omega, x)\right) = f(\omega, x)\ , \quad \text{in } D \subset \mathbb{R}^d, \quad \omega \in \Omega\ ,$$

with random coefficient functions $a(\omega, \cdot)\colon D \to \mathbb{R}$ and $f(\omega, \cdot)\colon D \to \mathbb{R}$, equipped with suitable boundary conditions. Similarly to the SDE problem above, the lower the regularity of the random coefficients and/or the functional φ, the lower the variance decay rate β becomes, cf. [29]. Moreover, the computational cost rate γ is typically proportional to the dimension d of the spatial domain D.

Finally, let us mention that MLMC has been successfully applied to a wide range applications, such as seismic wave propagation [3], stochastic reaction networks [2, 27], stochastic partial differential equations [5, 26], optimal experimental design [7], Markov chain Monte Carlo simulation [11, 20], Bayesian inversion and filtering methods [21, 8, 25, 16], and rare event estimation/importance sampling [30, 22], to name but a few. As a consequence of these applications' diverse nature, a wide variety of different rate triplet scenarios is commonly relevant in practice.

2. Theory

In this section we derive weak assumptions under which the normalized MLMC estimator $(\mathcal{A}_{ML} - \mathbb{E}[X_L])/\sqrt{\operatorname{Var}(\mathcal{A}_{ML})}$ converges in distribution to a standard normal as $\epsilon \to 0$. The main tool used for verifying the CLT will be the Lindeberg condition, which in its classical formulation is an integrability condition for triangular arrays of independent random variables (r.v.) Y_{nm}, with $n \in \mathbb{N}$ and $1 \leq m \leq k_n$; cf. [12]. However, in the multilevel setting it is more convenient to work with generalized triangular arrays of independent r.v. of the form $Y_{\epsilon m}$, which for a fixed $\epsilon > 0$ take possible non-zero elements within the set of indices $1 \leq m \leq n(\epsilon)$, where $n\colon (0, \infty) \to \mathbb{N}$ is a strictly decreasing function of $\epsilon > 0$ with $\lim_{\epsilon \downarrow 0} n(\epsilon) = \infty$.

The following theorem is a trivial extension of [23] from triangular arrays to generalized triangular arrays.

Theorem 2.1 (Lindeberg-Feller Theorem)**.** *For every $\epsilon > 0$, let $\{Y_{\epsilon m}\}$, $1 \leq m \leq n(\epsilon)$ with $n\colon (0, \infty) \to \mathbb{N}$ and $\lim_{\epsilon \downarrow 0} n(\epsilon) = \infty$ be a generalized triangular array of independent random variables that are centered and nor-*

malized, so that

$$\mathbb{E}[Y_{\epsilon m}] = 0 \quad \textit{and} \quad \sum_{m=1}^{n(\epsilon)} \mathbb{E}\left[Y_{\epsilon m}^2\right] = 1 \ , \tag{15}$$

respectively. Then, the Lindeberg condition:

$$\lim_{\epsilon \downarrow 0} \sum_{m=1}^{n(\epsilon)} \mathbb{E}\left[Y_{\epsilon m}^2 \mathbf{1}_{\{|Y_{\epsilon m}|>\nu\}}\right] = 0 \quad \forall\, \nu > 0 \ , \tag{16}$$

holds, if and only if

$$\sum_{m=1}^{n(\epsilon)} Y_{\epsilon m} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \ \textit{as}\ \epsilon \downarrow 0 \quad \textit{and} \quad \lim_{\epsilon \downarrow 0} \max_{m \in \{1,2,\dots,n(\epsilon)\}} \mathbb{E}\left[Y_{\epsilon m}^2\right] = 0 \ . \tag{17}$$

We will refer to (17) as the *extended CLT condition.* By defining

$$n(\epsilon) := \sum_{\ell=0}^{L} M_\ell, \tag{18}$$

and

$$Y_{\epsilon m} := \begin{cases} \frac{\Delta_0 X^m - \mathbb{E}[\Delta_0 X]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})} M_0} & m \leq M_0 \\ \frac{\Delta_1 X^m - \mathbb{E}[\Delta_1 X]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})} M_1} & M_0 < m \leq M_0 + M_1 \\ \vdots & \\ \frac{\Delta_L X^m - \mathbb{E}[\Delta_L X]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})} M_L} & n(\epsilon) - M_L < m \leq n(\epsilon), \end{cases} \tag{19}$$

the normalized variance minimizing MLMC estimator can be represented by generalized triangular arrays as follows:

$$\frac{\mathcal{A}_{ML} - \mathbb{E}[X_L]}{\sqrt{\mathrm{Var}(\mathcal{A}_{ML})}} = \sum_{m=1}^{n(\epsilon)} Y_{\epsilon m} \ . \tag{20}$$

We note that the telescoping property $\mathbb{E}[X_L] = \sum_{\ell=0}^{L} \mathbb{E}[\Delta_\ell X]$ was used to obtain (20). Moreover, the representation (20) and the below corollary trivially extends to any normalized MLMC estimator.

Corollary 2.2. *Let* $\mathcal{A}_{ML}$ *denote the variance minimizing MLMC estimator applied to estimate the expectation of* $X \in L^2(\Omega)$ *based on the collection of r.v.* $\{X_\ell\} \subset L^2(\Omega)$ *satisfying Assumption 1.1. Suppose that* $\mathrm{Var}(\mathcal{A}_{ML}) > 0$

for any $\epsilon > 0$. Then the normalized estimator (20) *satisfies the extended CLT condition* (17)*, if and only if for any $\nu > 0$,*

$$\lim_{\epsilon \downarrow 0} \sum_{\ell=0}^{L} \frac{V_\ell}{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell^2}{V_\ell} \nu \right\}} \right] = 0. \tag{21}$$

Proof. For all $\epsilon > 0$, the triangular array representation (20) of the MLMC estimator obviously satisfies the centering and normalization conditions (15), and its elements are centered and mutually independent. By Theorem 2.1, the extended CLT condition thus holds if and only if Lindeberg's condition (16) holds. For any $\nu > 0$, here Lindeberg's condition takes the form:

$$\begin{aligned}
&\lim_{\epsilon \to 0} \sum_{m=1}^{n(\epsilon)} \mathbb{E}\left[Y_{\epsilon m}^2 \mathbf{1}_{\{|Y_{\epsilon m}| > \nu\}} \right] \\
&= \lim_{\epsilon \to 0} \sum_{\ell=0}^{L} \sum_{i=1}^{M_\ell} \mathbb{E}\left[\frac{\left|\Delta_\ell X^i - \mathbb{E}[\Delta_\ell X]\right|^2}{M_\ell^2 \mathrm{Var}(\mathcal{A}_{ML})} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X^i - \mathbb{E}[\Delta_\ell X]|^2}{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell^2} > \nu^2 \right\}} \right] \\
&= \lim_{\epsilon \downarrow 0} \sum_{\ell=0}^{L} \frac{V_\ell}{M_\ell \mathrm{Var}(\mathcal{A}_{ML})} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell^2}{V_\ell} \nu^2 \right\}} \right].
\end{aligned}$$

□

Assumption 1.1 does not provide any lower bound on the decay rate of the variance sequence $\{V_\ell\}$, and therefore it alone is not sufficiently strong to ensure that Lindeberg's condition (21) holds in general. The problem is that without any lower bound on V_ℓ, there are asymptotic settings where a non-negligible contribution to the variance of the variance minimizing MLMC estimator derives from a finite number of samples.

Example 2.1. Consider the setting where $\beta \leq 2\alpha < \gamma$, for some constants $c_2 > c_1 > 0$,

$$c_1 2^{-2\alpha\ell} \leq V_\ell \leq c_2 2^{-\beta\ell} \qquad \forall \ell \in \mathbb{N}_0,$$

and for an infinite subsequence $\{k_i\} \subset \mathbb{N}_0$,

$$V_{k_i} = \Theta_i(2^{-2\alpha k_i}) \quad \text{and} \quad S_{k_i} = \Theta_i(2^{(\gamma-2\alpha)k_i/2}) \quad \forall i \in \mathbb{N}_0.$$

Then equation (2) implies there exists $c, C, \tilde{c}, \hat{c} \in \mathbb{R}_+$ such that for all $y \in \{\epsilon > 0 \mid L(\epsilon) \in \{k_i\}\}$,

$$1 \leq M_{L(y)} < C,$$

and

$$\hat{c} \leq \max\left(\frac{V_{L(y)}}{M_{L(y)}\mathrm{Var}(\mathcal{A}_{ML}(y))}, \frac{M_{L(y)}^2\mathrm{Var}(\mathcal{A}_{ML}(y))}{V_{L(y)}}\right) \leq \tilde{c}.$$

Hence, for any $\nu < (2\tilde{c})^{-1}$,

$$\begin{aligned}
&\limsup_{\epsilon\downarrow 0} \sum_{\ell=0}^{L} \frac{V_\ell}{M_\ell \mathrm{Var}(\mathcal{A}_{ML})} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\mathrm{Var}(\mathcal{A}_{ML})M_\ell^2}{V_\ell}\nu\right\}}\right] \\
&\geq \limsup_{\epsilon\downarrow 0} \frac{V_L}{M_L \mathrm{Var}(\mathcal{A}_{ML})} \mathbb{E}\left[\frac{|\Delta_L X - \mathbb{E}[\Delta_L X]|^2}{V_L} \mathbf{1}_{\left\{\frac{|\Delta_L X - \mathbb{E}[\Delta_L X]|^2}{V_L} > \frac{\mathrm{Var}(\mathcal{A}_{ML})M_L^2}{V_L}\nu\right\}}\right] \\
&\geq \limsup_{i\to\infty} \hat{c}\, \mathbb{E}\left[\frac{|\Delta_{k_i} X - \mathbb{E}[\Delta_{k_i} X]|^2}{V_{k_i}} \mathbf{1}_{\left\{\frac{|\Delta_{k_i} X - \mathbb{E}[\Delta_{k_i} X]|^2}{V_{k_i}} > \frac{1}{2}\right\}}\right] \geq \frac{\hat{c}}{2} > 0.
\end{aligned}$$

Example 2.1 illustrates that Assumption 1.1 is not sufficiently strong to ensure condition (21) when $\gamma > \beta$. We therefore impose the following additional variance decay assumption, which can be viewed as an implicit weak lower bound on the sequence $\{V_\ell\}$.

Assumption 2.1. If Assumption 1.1 holds for a collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ with limit $X \in L^2(\Omega)$ in the setting $\gamma > \beta$ and $\lim_{\ell\to\infty} S_\ell = \infty$, then assume additionally that $\beta < 2\alpha$ and that there exists an $\upsilon \in [\beta, 2\alpha)$ such that

$$\liminf_{\ell\to\infty} S_\ell 2^{(\upsilon-\gamma)\ell/2} > 1.$$

Lemma 2.3. *Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumptions 1.1 and 2.1. Then*

$$\lim_{\epsilon\downarrow 0} \frac{\mathrm{Var}(\mathcal{A}_{ML})}{\epsilon^2} = 1\ . \tag{22}$$

Proof. For any $\epsilon > 0$, it follows from equation (2) that

$$\frac{\mathrm{Var}(\mathcal{A}_{ML})}{\epsilon^2} = \sum_{\ell=0}^{L} \frac{V_\ell}{\epsilon^2 M_\ell} \leq \sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} = 1\ ,$$

and by the mean value theorem there exists a constant $C > 0$ such that

$$\begin{aligned} \sum_{\ell=0}^{L} \frac{V_\ell}{\epsilon^2 M_\ell} &\geq \sum_{\ell=0}^{L} \mathbf{1}_{\{V_\ell > 0\}} \frac{V_\ell}{\sqrt{\frac{V_\ell}{C_\ell}} S_L + \epsilon^2} \\ &\geq 1 - \sum_{\ell=0}^{L} \mathbf{1}_{\{V_\ell > 0\}} \frac{V_\ell \epsilon^2}{\frac{V_\ell}{C_\ell} S_L^2} \\ &\geq 1 - \epsilon^2 \frac{\sum_{\ell=0}^{L} C_\ell}{S_L^2} \\ &\geq 1 - C\epsilon^2 \frac{2^{\gamma L}}{S_L^2} \ . \end{aligned} \tag{23}$$

To complete the proof, it remains to verify that

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon^2 2^{\gamma L}}{S_L^2} = 0 \ . \tag{24}$$

We separate the proof into three cases:

(i): If $\beta < \gamma$ and $\lim_{\ell \to \infty} S_\ell = \infty$, then Assumption 2.1 implies that

$$\frac{\epsilon^2 2^{\gamma L}}{S_L^2} = \mathcal{O}(\epsilon^{2 - \upsilon/\alpha}) \ ,$$

and since $\upsilon < 2\alpha$, the claim follows.

(ii): If $\beta = \gamma$ and $\lim_{\ell \to \infty} S_\ell = \infty$, then $\gamma \leq 2\alpha$, cf. Assumption 1.1, implies that $\epsilon^2 2^{\gamma L} = \mathcal{O}_\epsilon(1)$ and the claim follows.

(iii): If $\lim_{\ell \to \infty} S_\ell =: S < \infty$, then there exists a $k > 1$ such that $\gamma/k < 2\alpha$ and a $C > 0$ such that

$$\frac{\mathrm{Var}(\mathcal{A}_{ML})}{\epsilon^2} \geq \sum_{\ell=0}^{\lceil L/k \rceil} \mathbf{1}_{\{V_\ell > 0\}} \frac{V_\ell}{\sqrt{\frac{V_\ell}{C_\ell}} S + \epsilon^2} \geq \frac{S_{\lceil L/k \rceil}}{S} - C\epsilon^2 \frac{2^{\gamma L/k}}{S^2},$$

The claim follows from $\lim_{\epsilon \downarrow 0} \epsilon^2 2^{\gamma L/k} = 0$ and $\lim_{\epsilon \downarrow 0} S_{\lceil L/k \rceil} = S$.

Case **(iii)** covers all settings $\gamma \geq \beta$ which are not covered by either **(i)** or **(ii)**. Furthermore, since $S_\ell = \mathcal{O}_\ell(2^{(\gamma - \beta)\ell/2})$, it is clear that **(iii)** also covers all settings with $\beta > \gamma$. This shows that cases **(i)**–**(iii)** cover all settings that are valid under Assumptions 1.1 and 2.1.

□

Lemma 2.3 implies that we can reformulate Lindeberg's condition for the MLMC estimator as follows:

Corollary 2.4. *Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumptions 1.1 and 2.1. Then the normalized MLMC estimator satisfies the extended CLT condition* (17)*, if and only if for any $\nu > 0$,*

$$\lim_{\epsilon \downarrow 0} \sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbf{1}_{\{V_\ell > 0\}} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{V_\ell} \nu \right\}} \right] = 0 \,. \tag{25}$$

Proof. From the proof of Lemma 2.3 it follows that there exists an $\bar{\epsilon} > 0$ such that

$$\frac{1}{2} \leq \frac{\mathrm{Var}(\mathcal{A}_{ML})}{\epsilon^2} \leq 1 \,, \quad \forall \epsilon \in (0, \bar{\epsilon}) \,.$$

Consequently, for any $\epsilon \in (0, \bar{\epsilon})$ and any $\nu > 0$ we have that

$$\begin{aligned} &\sum_{\ell=0}^{L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell^2}{V_\ell} \nu \right\}} \right] \\ &\geq \sum_{\ell=0}^{L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{\epsilon^2 M_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{V_\ell} \nu \right\}} \right] , \end{aligned}$$

as well as

$$\begin{aligned} &\sum_{\ell=0}^{L} \frac{1}{\mathrm{Var}(\mathcal{A}_{ML})} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{M_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\mathrm{Var}(\mathcal{A}_{ML}) M_\ell^2}{V_\ell} \nu \right\}} \right] \\ &\leq 2 \sum_{\ell=0}^{L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{\epsilon^2 M_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{2 V_\ell} \nu \right\}} \right] . \end{aligned}$$

These upper and lower bounds imply that that Lindeberg's condition (21) is equivalent to the following condition: for any $\nu > 0$ it holds that

$$\lim_{\epsilon \downarrow 0} \sum_{\ell=0}^{L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{\epsilon^2 M_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{V_\ell} \nu \right\}} \right] = 0 \,.$$

Following similar steps as those leading to inequality (23), we further note that for sufficiently small $\epsilon > 0$,

$$\sum_{\ell=0}^{L} \frac{1}{\epsilon^2 M_\ell} \mathbb{E}\left[|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2 \mathbf{1}_{\left\{\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{\epsilon^2 M_\ell^2} > \nu\right\}} \right]$$
$$= \sum_{\ell=0}^{L} \left\{ \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{V_\ell}\nu\right\}} \right] \right\} - \rho(\epsilon) \; , \tag{26}$$

where the mapping $\rho\colon \mathbb{R}_+ \to [0,\infty)$, satisfying $\lim_{\epsilon\downarrow 0} \rho(\epsilon) = 0$, can be derived as in the proof of Lemma 2.3. □

In settings with $\lim_{\ell\to\infty} S_\ell < \infty$, the summability of the sequence $\{\sqrt{C_\ell V_\ell}\}$ turns out to be sufficient to prove that the extended CLT condition holds.

Theorem 2.5. *Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to estimate the expectation of $X \in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\} \subset L^2(\Omega)$ satisfying Assumptions 1.1 and $\lim_{\ell\to\infty} S_\ell < \infty$. Then the extended CLT condition* (17) *is satisfied for the normalized estimator.*

Note that the setting $\beta > \gamma$ is completely covered by Theorem 2.5, as then

$$S := \lim_{k\to\infty} S_k = \lim_{k\to\infty} \sum_{\ell=0}^{k} \sqrt{V_\ell C_\ell} \leq c \lim_{k\to\infty} \sum_{\ell=0}^{k} 2^{(\gamma-\beta)\ell/2} < \infty \; .$$

Proof. We prove this result by verifying that condition (25) holds.

As the sequence $\{S_\ell\}$ is monotonically increasing, it is contained in the bounded interval $[S_0, S]$ with $S_0 > 0$. Consequently, Lindeberg's condition (25) is equivalent to:

$$\lim_{\epsilon\downarrow 0} \sum_{\ell=0}^{L} \mathbf{1}_{\{V_\ell>0\}} \sqrt{\frac{C_\ell}{V_\ell}} \mathbb{E}\left[|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2 \mathbf{1}_{\{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2 > \epsilon^2 M_\ell^2 \nu\}} \right] = 0 \; , \quad \forall\, \nu > 0 \; .$$

Fix a $\nu > 0$. Then for all $\ell \in \mathbb{N}_0$,

$$\mathbb{E}\left[|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2 \mathbf{1}_{\{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2 > \epsilon^2 M_\ell^2 \nu\}} \right] \leq V_\ell.$$

By the preceding inequality and the summability of the sequence $\{V_\ell C_\ell\}$, the dominated convergence theorem yields that

$$
\begin{aligned}
&\lim_{\epsilon\downarrow 0}\sum_{\ell=0}^{L}\mathbf{1}_{\{V_\ell>0\}}\sqrt{\frac{C_\ell}{V_\ell}}\mathbb{E}\left[\left|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]\right|^2\mathbf{1}_{\left\{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2>\epsilon^2 M_\ell^2\nu\right\}}\right]\\
&=\sum_{\ell=0}^{\infty}\mathbf{1}_{\{V_\ell>0\}}\sqrt{\frac{C_\ell}{V_\ell}}\lim_{\epsilon\downarrow 0}\mathbb{E}\left[\left|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]\right|^2\mathbf{1}_{\left\{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2>\epsilon^2 M_\ell^2\nu\right\}}\right].
\end{aligned}
\tag{27}
$$

For all $\ell\in\mathbb{N}_0$ such that $V_\ell>0$,

$$
\lim_{\epsilon\downarrow 0}\epsilon^2 M_\ell^2(\epsilon)\geq\lim_{\epsilon\downarrow 0}\epsilon^{-2}\frac{V_\ell}{C_\ell}S_L^2=\infty,
$$

and the dominated convergence theorem applies for all $\ell\in\mathbb{N}_0$:

$$
\begin{aligned}
&\mathbf{1}_{\{V_\ell>0\}}\sqrt{\frac{C_\ell}{V_\ell}}\lim_{\epsilon\downarrow 0}\mathbb{E}\left[\left|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]\right|^2\mathbf{1}_{\left\{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2>\epsilon^2 M_\ell^2\nu\right\}}\right]\\
&=\mathbf{1}_{\{V_\ell>0\}}\sqrt{\frac{C_\ell}{V_\ell}}\mathbb{E}\left[\lim_{\epsilon\downarrow 0}\left|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]\right|^2\mathbf{1}_{\left\{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2>\epsilon^2 M_\ell^2\nu\right\}}\right]\\
&=0.
\end{aligned}
\tag{28}
$$

As the above argument is valid for any fixed $\nu>0$, equations (27) and (28) verify that Lindeberg's condition holds.

□

We next verify the extended CLT condition for the variance minimizing MLMC estimator in settings with $\lim_{\ell\to\infty}S_\ell=\infty$.

Theorem 2.6. *Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to estimate the expectation of $X\in L^2(\Omega)$ based on the collection of r.v. $\{X_\ell\}\subset L^2(\Omega)$ satisfying Assumptions 1.1 and 2.1. Assume that $\lim_{\ell\to\infty}S_\ell=\infty$ and that*

$$
\lim_{\ell\to\infty}\mathbf{1}_{\{V_\ell>0\}}\mathbb{E}\left[\frac{\left|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]\right|^2}{V_\ell}\mathbf{1}_{\left\{\frac{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2}{V_\ell}>2^{(2\alpha-\gamma)\ell}S_\ell^2\nu\right\}}\right]=0
$$

holds for any $\nu>0$. Then the extended CLT condition (17) *is satisfied for the normalized MLMC estimator.*

Proof. From (2) and $C_\ell = \Theta_\ell(2^{\gamma\ell})$ it follows that there exists a $c > 0$ such that

$$\frac{\epsilon^2 M_\ell^2}{V_\ell} \geq \frac{\epsilon^{-2} S_\ell^2}{C_\ell} > c2^{(2\alpha-\gamma)\ell} S_\ell^2.$$

Consequently,

$$\sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \frac{\epsilon^2 M_\ell^2}{V_\ell} \nu \right\}} \right]$$
$$\leq \sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \nu c 2^{(2\alpha-\gamma)\ell} S_\ell^2 \right\}} \right].$$

Let $\widetilde{L} : (0, \infty) \to \mathbb{N}_0$ be a monotonically decreasing function satisfying the constraints

$$\lim_{\epsilon \downarrow 0} \widetilde{L}(\epsilon) = \infty \quad \text{and} \quad \lim_{\epsilon \downarrow 0} \frac{S_{\widetilde{L}(\epsilon)}}{S_{L(\epsilon)}} = 0.$$

Under the current assumption $\lim_{\epsilon \downarrow 0} S_{L(\epsilon)} = \infty$, it is always possible to construct such an $\widetilde{L}$, e.g.,

$$\widetilde{L}(\epsilon) := \min\left\{ \ell \in \mathbb{N}_0 \mid S_{\ell+1} \geq \sqrt{S_{L(\epsilon)}} \right\}.$$

Provided that $\epsilon > 0$ is sufficiently small, it holds that $\widetilde{L} < L$ and we may write

$$\sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \nu c 2^{(2\alpha-\gamma)\ell} S_\ell^2 \right\}} \right]$$
$$\leq \sum_{\ell=0}^{\widetilde{L}} \frac{\sqrt{V_\ell C_\ell}}{S_L} + \sum_{\ell=\widetilde{L}+1}^{L} \frac{\sqrt{V_\ell C_\ell}}{S_L} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \nu c 2^{(2\alpha-\gamma)\ell} S_\ell^2 \right\}} \right]$$
$$\leq \frac{S_{\widetilde{L}}}{S_L} + \frac{S_L - S_{\widetilde{L}}}{S_L} \times \sup_{\ell > \widetilde{L}} \mathbb{E}\left[\frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} \mathbf{1}_{\left\{ \frac{|\Delta_\ell X - \mathbb{E}[\Delta_\ell X]|^2}{V_\ell} > \nu c 2^{(2\alpha-\gamma)\ell} S_\ell^2 \right\}} \right].$$

Consequently,

$$
\begin{aligned}
&\lim_{\epsilon\downarrow 0}\sum_{\ell=0}^{L}\frac{\sqrt{V_\ell C_\ell}}{S_L}\mathbb{E}\left[\frac{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2}{V_\ell}\mathbf{1}_{\left\{\frac{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2}{V_\ell}>\frac{\epsilon^2 M_\ell^2}{V_\ell}\nu\right\}}\right]\\
&\leq \lim_{\epsilon\downarrow 0}\frac{S_{\widetilde{L}}}{S_L}+\limsup_{\ell\to\infty}\mathbb{E}\left[\frac{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2}{V_\ell}\mathbf{1}_{\left\{\frac{|\Delta_\ell X-\mathbb{E}[\Delta_\ell X]|^2}{V_\ell}>\nu c 2^{(2\alpha-\gamma)\ell}S_\ell^2\right\}}\right]\\
&=0.
\end{aligned}
$$

□

2.1. CLT for the mass-shifted MLMC estimator

The key feature of the mass-shifted MLMC estimator that is particularly handy for proving the CLT is that irrespective of whether $\{S_\ell\}$ is uniformly bounded from above or not, it will always be the case that $\lim_{\ell\to\infty}\widetilde{S}_\ell<\infty$. The CLT follows by this property and an extension of Theorem 2.5.

Proof of Theorem 1.2. Recall that the mass-shifted MLMC estimator is given by

$$\widetilde{\mathcal{A}}_{ML}=\sum_{\ell=0}^{L}\sum_{i=1}^{\widetilde{M}_\ell}\frac{\Delta_\ell X^i}{\widetilde{M}_\ell},$$

where $\widetilde{M}_\ell$ for a given $\xi>0$ is defined in equation (9) and $\{\Delta_\ell X\}$ is a sequence of r.v. satisfying Assumption 1.1 for a rate triplet α,β,γ. Let $\{Y_\ell\}_{\ell=-1}^{\infty}\subset L^2(\Omega)$ denote a auxiliary sequence satisfying $Y_{-1}:=0$ and for all $\ell\geq 0$,

$$Y_\ell\overset{\mathrm{d}}{=}X_\ell,\qquad \Delta_\ell Y\overset{\mathrm{d}}{=}\Delta_\ell X,$$

and

$$\overline{C}_\ell:=\mathrm{Cost}(\Delta_\ell Y)=\frac{\mathrm{Cost}(\Delta_\ell X)}{(S_\ell+1)^2\log(S_\ell+1)^{2(1+\xi)}}=\frac{C_\ell}{(S_\ell+1)^2\log(S_\ell+1)^{2(1+\xi)}}.$$

Let $\mathcal{A}_{ML}$ denote the variance minimizing MLMC estimator applied to $\{Y_\ell\}_{\ell=-1}^{\infty}$, i.e.,

$$\mathcal{A}_{ML}=\sum_{\ell=0}^{L}\sum_{i=1}^{M_\ell}\frac{\Delta_\ell Y^i}{M_\ell},\tag{29}$$

where it follows by $\mathrm{Var}(\Delta_\ell Y)=\mathrm{Var}(\Delta_\ell X)=V_\ell$ and equation (2) that

$$M_\ell=\max\left(\left\lceil\epsilon^{-2}\sqrt{\frac{V_\ell}{\overline{C}_\ell}}\sum_{\ell=0}^{L}\sqrt{V_\ell\overline{C}_\ell}\right\rceil,1\right).$$

By construction,

$$\sum_{\ell=0}^{L} \sqrt{V_\ell \overline{C}_\ell} = \widetilde{S}_L,$$

hence, $M_\ell = \widetilde{M_\ell}$ for all $\ell \in [0, L]$. Consequently, $\mathcal{A}_{ML} \stackrel{\mathrm{d}}{=} \widetilde{\mathcal{A}}_{ML}$, so the theorem follows if we can prove the CLT for the normalized version of $\mathcal{A}_{ML}$.

The collection of random variables $\{Y_\ell\}$ satisfies the following slightly altered version of Assumption 1.1 (where $\Theta_\ell(2^{\gamma\ell})$ is replaced by $\mathcal{O}_\ell(2^{\bar{\gamma}\ell})$ in condition (iii)):

(i) for some $c_\alpha > 0$, $|\mathbb{E}[X - Y_\ell]| \leq c_\alpha 2^{-\alpha\ell}$ for all $\ell \geq 0$,

(ii) $\mathrm{Var}(\Delta_\ell Y) = \mathcal{O}_\ell(2^{-\beta\ell})$,

(iii) $\overline{C}_\ell = \mathcal{O}_\ell(2^{\bar{\gamma}\ell})$ and $\inf_{\ell \in \mathbb{N}_0} \overline{C}_\ell > c > 0$,

where $\bar{\gamma} \in (0, \gamma]$ and $\alpha, \beta, \gamma > 0$ (as everywhere else in this proof) stems from the rate triplet of $\{X_\ell\}$. Moreover,

$$\min(\beta, \gamma) \leq 2\alpha \implies \min(\beta, \bar{\gamma}) \leq 2\alpha,$$

and since $\{S_\ell\}$ is monotonically increasing,

$$\begin{aligned}
\widetilde{S}_L &= \sum_{\ell=0}^{L} \frac{\sqrt{V_\ell C_\ell}}{(S_\ell + 1)\log(S_\ell + 1)^{1+\xi}} \\
&= \sum_{\ell=0}^{L} \frac{S_\ell - S_{\ell-1}}{(S_\ell + 1)\log(S_\ell + 1)^{1+\xi}} \\
&\leq \int_{S_0}^{S_L} \frac{1}{(s+1)\log(s+1)^{1+\xi}} ds \\
&< \frac{1}{\xi \log(S_0 + 1)^\xi} < \infty.
\end{aligned}$$

This shows that $\widetilde{S}_\ell \in [\widetilde{S}_0, \widetilde{S}]$ for all $\ell \geq 0$, where $\widetilde{S}_0 = V_0 \overline{C}_0 > 0$ and $\widetilde{S} = \lim_{\ell \to \infty} \widetilde{S}_\ell < \infty$. Using the uniform bounds on $\{\widetilde{S}_\ell\}$ and the properties of the rate triplet for $\{Y_\ell\}$, the proofs of Lemma 2.3, Corollary 2.4 and Theorem 2.5 straightforwardly extends to the current setting, verifying the CLT for the normalized version of the estimator (29). □

References

[1] M. Ben Alaya, A. Kebaier. Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.* 25 (1) (2015) 211–234.

[2] D. F. Anderson and D. J. Higham. Multilevel monte carlo for continuous time markov chains, with applications in biochemical kinetics. *Multiscale Modeling & Simulation*, 10(1):146–179, 2012.

[3] M. Ballesio, J. Beck, A. Pandey, L. Parisi, E. von Schwerin, R. Tempone. Multilevel Monte Carlo Acceleration of Seismic Wave Propagation under Uncertainty. *arXiv:1810.01710*, 2018..

[4] A. Barth and A. Lang. Multilevel monte carlo method with applications to stochastic partial differential equations. *International Journal of Computer Mathematics*, 89(18):2479–2498, 2012.

[5] A. Barth, A. Lang, and C. Schwab. Multilevel monte carlo method for parabolic stochastic partial differential equations. *BIT Numerical Mathematics*, 53(1):3–27, 2013.

[6] C. Bayer, P. K. Friz, S. Riedel, J. Schoenmakers. From rough path estimates to multilevel Monte Carlo. *SIAM J. Numer. Anal.* 54 (3) (2016) 1449–1483.

[7] J. Beck, B. M. Dia, L. FR Espath, R. Tempone. Multilevel Double Loop Monte Carlo and Stochastic Collocation Methods with Importance Sampling for Bayesian Optimal Experimental Design. *arXiv:1811.11469*, 2018.

[8] A. Chernov, H. Hoel, K. JH Law, F. Nobile, and R. Tempone. Multilevel ensemble kalman filtering for spatially extended models. *arXiv:1608.08558*, 2016.

[9] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone. A continuation multilevel Monte Carlo algorithm. *BIT* 55 (2) (2015) 399–432.

[10] S. Dereich, S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. *Ann. Appl. Probab.* 26 (1) (2016) 136–185.

[11] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A hierarchical multilevel markov chain monte carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, 2015.

[12] R. Durrett. Probability: theory and examples. 2nd Edition, Duxbury Press, Belmont, CA, 1996.

[13] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[14] M B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.

[15] D. Giorgi, V. Lemaire, G. Pagès. Limit theorems for weighted and regular multilevel estimators. *Monte Carlo Methods Appl.* 23 (1) (2017) 43–70.

[16] A. Gregory, C. J. Cotter, and S. Reich. Multilevel ensemble transform particle filtering. *SIAM Journal on Scientific Computing*, 38(3):A1317–A1338, 2016.

[17] H. Hoel, J. Häppölä, and R. Tempone. Construction of a mean square error adaptive euler–maruyama method with applications in multilevel Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo Methods*, p. 29–86. Springer, 2016.

[18] H. Hoel, E. Von Schwerin, A. Szepessy, R. Tempone. Implementation and analysis of an adaptive multilevel monte carlo algorithm. *Monte Carlo Methods and Applications*, 20(1):1–41, 2014.

[19] S. Heinrich. Monte Carlo complexity of global solution of integral equations. *J. Complexity*, 14(2):151–175, 1998.

[20] V. H. Hoang, C. Schwab, and A. M. Stuart. Complexity analysis of accelerated mcmc methods for bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[21] A. Jasra, K. Kamatani, K. JH Law, and Y. Zhou. Multilevel particle filters. *SIAM Journal on Numerical Analysis*, 55(6):3068–3096, 2017.

[22] A. Kebaier and J. Lelong. Coupling importance sampling and multilevel monte carlo using sample average approximation. *Methodology and Computing in Applied Probability*, 20(2):611–641, 2018.

[23] A. Klenke. Probability theory. 2nd Edition, *Universitext*, Springer, London, 2014.

[24] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[25] J. Latz, I. Papaioannou, and E. Ullmann. Multilevel sequential monte carlo for bayesian inverse problems. *Journal of Computational Physics*, 368:154–178, 2018.

[26] S. Mishra and C. Schwab. Monte-carlo finite-volume methods in uncertainty quantification for hyperbolic conservation laws. In *Uncertainty Quantification for Hyperbolic and Kinetic Equations*, pages 231–277. Springer, 2017.

[27] A. Moraes, R. Tempone, and P. Vilanova. Multilevel hybrid chernoff tau-leap. *BIT Numerical Mathematics*, 56(1):189–239, 2016.

[28] C.-han Rhee and P. W. Glynn. Unbiased estimation with square root convergence for sde models. *Operations Research*, 63(5):1026–1043, 2015.

[29] A. L. Teckentrup, R. Scheichl, M. B. Giles, E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients, *Numer. Math.* 125 (3) (2013) 569–600.

[30] E. Ullmann and I. Papaioannou. Multilevel estimation of rare events. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):922–953, 2015.

[31] Z. Zheng, J. Blanchet, and P. W. Glynn. Rates of convergence and clts for subcanonical debiased mlmc. In *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 465–479. Springer, 2016.

[32] Z. Zheng and P. W. Glynn. A CLT for infinitely stratified estimators, with applications to debiased mlmc. *ESAIM: Proceedings and Surveys*, 59:104–114, 2017.