Rates of Convergence and CLTs for Subcanonical Debiased MLMC

Zeyu Zheng, Jose Blanchet and Peter W. Glynn

Abstract In constructing debiased multi-level Monte Carlo (MLMC) estimators, one must choose a randomization distribution. In some algorithmic contexts, an optimal choice for the randomization distribution leads to a setting in which the mean time to generate an unbiased observation is infinite. This paper extends the well known efficiency theory for Monte Carlo algorithms in the setting of a finite mean for this generation time to the infinite mean case. The theory draws upon stable law weak convergence results, and leads directly to exact convergence rates and central limit theorems (CLTs) for various debiased MLMC algorithms, most particularly as they arise in the context of stochastic differential equations. Our CLT theory also allows simulators to construct asymptotically valid confidence intervals for such infinite mean MLMC algorithms.

Keywords Monte Carlo estimator efficiency · Central limit theorems
Subcanonical convergence rates · Infinite mean generation time · Infinite variance

1 Introduction

In comparing Monte Carlo algorithms, a key result in the literature concerns the efficiency trade-off between the variance of an estimator, and the computer time required to compute that estimator. In particular, suppose that a quantity $z = E(X)$ is to be computed. The associated Monte Carlo estimator is constructed by generating independent and identically distributed (iid) copies $X_1, X_2, \ldots$ of X; the computer time required to generate X_i is given by τ_i, a positive random variable (rv). The

Z. Zheng (✉) · J. Blanchet · P. W. Glynn
Stanford University, Stanford, CA 94305, USA
e-mail: zyzheng@stanford.edu

J. Blanchet
e-mail: jblanchet@stanford.edu

P. W. Glynn
e-mail: glynn@stanford.edu

A. B. Owen and P. W. Glynn (eds.), *Monte Carlo and Quasi-Monte Carlo Methods*, Springer Proceedings in Mathematics & Statistics 241,
https://doi.org/10.1007/978-3-319-91436-7_26

(X_i, τ_i) pairs are then iid in i, where X_i and τ_i are generally correlated. Given a computer time budget c, let $N(c)$ be the number of X_i's generated in c units of computer time, so that $N(c) = \max\{n \geq 0 : \tau_1 + \cdots + \tau_n \leq c\}$. The estimator for z that is available with computational budget c is then $\bar{X}_{N(c)}$, where $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$. It is well known that when $E(\tau_1) < \infty$ and $\mathrm{Var}(X_1) < \infty$, the central limit theorem (CLT)

$$c^{1/2}\left(\bar{X}_{N(c)} - z\right) \Rightarrow \sqrt{E(\tau_1) \cdot \mathrm{Var}(X_1)}\, N(0, 1) \tag{1}$$

holds as $c \to \infty$, where $\Rightarrow$ denotes weak convergence, and $N(0, 1)$ denotes a normal rv with mean 0 and variance 1. With (1) in hand, one can now compare the efficiency of different algorithms (as associated with two rv's X and Y for which $E(X) = z = E(Y)$) for a given (large) computational budget c. The result (1) is discussed in [7], but is worked out in much greater detail in [5]. In the latter reference, the theory focuses on settings in which $E(\tau_1) < \infty$; an extension to $\mathrm{Var}(X_1) = \infty$ can also be found there.

In this paper, we extend this efficiency framework to the setting in which $E(\tau_1) = \infty$. As we shall argue in Sect. 3, this extension is useful in some applications of debiased MLMC; see [10]. In particular, there are various debiased MLMC algorithms which lead naturally to $E(\tau_1) = \infty$; such algorithms are believed to converge at a rate slower than the "canonical" $c^{-1/2}$ rate associated with (1), so that they exhibit "subcanonical rates." However, theoretical analysis of such algorithms has been hampered by the fact that no analog to (1) exists when $E(\tau_1) = \infty$. For example, much of the theory on subcanonical MLMC establishes upper bounds on the rate of convergence, but not lower bounds. Such lower bounds would follow automatically, in the presence of an analog to (1). Other references which study estimators based on Multilevel Monte Carlo (MLMC) via weak convergence methods include [1, 6, 9] (but they do not analyze debiased estimators, nor do they focus on the subcanonical case studied here).

It is worth noting that the act of terminating a debiased computation at a fixed computational budget inevitably introduces bias. This bias is theoretically inevitable, since any part of the sample space for the underlying random variables that takes more computation than provided by the budget can not be sampled within the given budget. Fortunately, the bias of the estimators discussed here typically goes to zero rapidly; see [4]. Furthermore, in the limit theorems described in this paper, the bias is always of smaller order than the sampling variability, as suggested by the fact that the limit random variables in all our theorems have mean zero.

This paper establishes limit theory for such subcanonical rate algorithms in Sect. 2 for both the case in which X has finite variance (Theorems 1 and 2) and when X is in the domain of attraction of a finite mean stable law (Theorem 3). Section 3 shows how the theory applies specifically to the debiased MLMC setting, and provides theory that slightly improves upon the known convergence rates for such algorithms in the stochastic differential equation context, and shows how the theory can be used to obtain asymptotically valid confidence intervals for such infinite mean procedures.

2 The Key Limit Theorems When $E(\tau_1) = \infty$

In the setting in which $E(\tau_1) = \infty$, limit theory for sums and averages typically fail to hold unless one makes strong assumptions about the tail behavior of τ_1. Consequently, we now require that τ_1 satisfy the tail condition:

A1. There exists $\alpha \in (0, 1]$ and a slowly varying function $L(\cdot)$ such that

$$P(\tau_1 > x) = x^{-\alpha} L(x)$$

for $x \geq 0$.

Remark 1 We note that a function $L(\cdot)$ is said to be *slowly varying* if for each $q > 0$, $L(qx)/L(x) \to 1$ as $x \to \infty$.

The assumption A1 is a strong requirement on the tail of τ_1 that comes close to asserting that τ_1 has a parametric-type Pareto tail. For typical Monte Carlo algorithms, there is no reason to believe that A1 will hold. However, in the debiased MLMC setting, the simulator must specify a randomization that strongly controls the distribution of τ_1. In this specific context, the randomization can be designed so that $\mathrm{Var}(X_1) < \infty$, with A1 describing the tail behavior of τ_1; see Sect. 3 for further discussion. (Requiring that $\mathrm{Var}(X_1) < \infty$ simplifies the construction of confidence intervals and the development of sequential procedures; see [10]).

In view of the above, we will focus first on the case where A1 holds with $\mathrm{Var}(X_1) < \infty$. The case in which $\alpha = 1$ is qualitatively different from the case in which $\alpha \in (0, 1)$. As it turns out, the most important applications of our theory in Sect. 3 concern the $\alpha = 1$ setting. Consequently, we start with this case. We assume here that $L(\cdot)$ takes the specific form

$$L(x) = a(\log x)^{\gamma} (\log \log x)^{\delta} \tag{2}$$

for $x \geq x_0$ and $a > 0$. If $\gamma < -1$ or if $\gamma = -1$ with $\delta < -1$, $E(\tau_1) < \infty$ and so this is covered by the theory presented in [5]. We therefore restrict our analysis to the case where $\gamma > -1$ or $\gamma = -1$ with $\delta \geq -1$.

Let $S_\alpha(\sigma, \beta, \mu)$ be a stable rv with index α, scale parameter σ, skewness parameter β, and shift parameter μ, with corresponding characteristic function

$$\begin{aligned} &E(\exp(i\theta S_\alpha(\sigma, \beta, \mu))) \\ &= \begin{cases} \exp\left(-\sigma^\alpha |\theta|^\alpha (1 - i\beta(\operatorname{sign}\theta)\tan(\pi\alpha/2)) + i\mu\theta\right), & \alpha \neq 1; \\ \exp\left(-\sigma |\theta| (1 + i\beta\frac{2}{\pi}(\operatorname{sign}\theta)\log(|\theta|)) + i\mu\theta\right), & \alpha = 1. \end{cases} \end{aligned}$$

Theorem 1 *Suppose $\sigma^2 = \mathrm{Var}(X_1) < \infty$. If $\alpha = 1$ and $L(\cdot)$ is as in (2), then*

$$\sqrt{\frac{c}{r(c)}} \left(\bar{X}_{N(c)} - z\right) \Rightarrow \sigma N(0, 1)$$

as $c \to \infty$, where

$$r(c) = \begin{cases} \frac{a}{1+\gamma}(\log c)^{1+\gamma}(\log\log c)^{\delta}, & \gamma > -1; \\ \frac{a}{1+\gamma}(\log\log c)^{1+\delta}, & \gamma = -1 < \delta; \\ a \log\log\log c, & \gamma = -1 = \delta. \end{cases}$$

Proof We start by noting that Theorem 4.5.1 of [11] implies that

$$\frac{\sum_{i=1}^{n} \tau_i - m_n}{c_n} \Rightarrow S_1(1, 1, 0) \tag{3}$$

as $n \to \infty$, where $(c_n : n \geq 1)$ is any sequence for which

$$\frac{nL(c_n)}{c_n} \to \frac{2}{\pi} \tag{4}$$

as $n \to \infty$, and $(m_n : n \geq 1)$ is chosen as

$$m_n = nc_n E(\sin(\tau_1/c_n)). \tag{5}$$

Given (2), (4) is satisfied by setting

$$c_n = \frac{\pi a}{2} n(\log n)^{\gamma}(\log\log n)^{\delta}.$$

As for m_n, fix $w > 0$ and write

$$m_n = nc_n \left(E(\sin(\tau_1/c_n)I(\tau_1 \leq wc_n)) + E(\sin(\tau_1/c_n)I(\tau_1 > wc_n))\right),$$

where $I(A)$ denotes the indicator function which is 1 when A occurs and 0 otherwise. Note that

$$nc_n \left|E(\sin(\tau_1/c_n)I(\tau_1 > wc_n))\right| \leq nc_n P(\tau_1 > wc_n) = O(c_n) \tag{6}$$

as $n \to \infty$, where $O(a_n)$ denotes any sequence for which $(|O(a_n)|/a_n : n \geq 1)$ is bounded.

On the other hand,

$$\begin{aligned} & nc_n E(\sin(\tau_1/c_n)I(\tau_1 \leq wc_n)) \\ &= nc_n E\left(\int_0^{\tau_1/c_n} \cos(y)dy I(\tau_1 \leq wc_n)\right) \\ &= nc_n \int_0^{w} \cos(y)P(yc_n < \tau_1 \leq wc_n)dy \\ &= nc_n \int_0^{w} \cos(y)P(\tau_1 > yc_n)dy - nc_n \sin(w)P(\tau_1 > wc_n) \end{aligned}$$

$$= nc_n \int_0^w \cos(y) P(\tau_1 > yc_n) dy + O(c_n) \tag{7}$$

as $n \to \infty$. But

$$\begin{aligned} n\cos(w) \int_0^{wc_n} P(\tau_1 > y) dy &\le nc_n \int_0^w \cos(y) P(\tau_1 > yc_n) dy \\ &\le n \int_0^{wc_n} P(\tau_1 > y) dy \end{aligned} \tag{8}$$

for $w \in [0, \pi/2]$. The upper and lower bounds in (8) follows from a change-of-variable arguement and the fact that $\cos(w) \le \cos(y)$ for any $0 \le y \le w \le \pi/2$. We shall argue below that

$$\int_0^{wc_n} P(\tau_1 > y) dy \sim r(c_n) \tag{9}$$

as $n \to \infty$, where we write $a_n \sim b_n$ as $n \to \infty$ whenever $a_n/b_n \to 1$ as $n \to \infty$. It is easily verified that $nr(c_n)/c_n \to \infty$, so it follows from (6), (7), and (8) that

$$\cos(w) \le \underline{\lim}_{n\to\infty} \frac{m_n}{nr(c_n)} \le \overline{\lim}_{n\to\infty} \frac{m_n}{nr(c_n)} \le 1. \tag{10}$$

By sending $w \to 0$ in (10), we conclude that

$$m_n \sim nr(c_n)$$

as $n \to \infty$.

Because $m_n/c_n \to \infty$, (3) implies that

$$\frac{1}{m_n} \sum_{i=1}^{n} \tau_i \Rightarrow 1$$

as $n \to \infty$, from which we find that

$$\frac{1}{c\eta} \sum_{i=1}^{\lfloor c\eta/r(c\eta) \rfloor} \tau_i \Rightarrow 1 \tag{11}$$

as $c \to \infty$ for any $\eta > 0$, where $\lfloor x \rfloor$ is the floor of x. If we choose $\eta = 1 + \varepsilon$ and $\eta = 1 - \varepsilon$ in (11), and use the fact that $r(\cdot)$ is slowly varying, we are led to the conclusion that

$$\frac{N(c)r(c)}{c} \Rightarrow 1 \tag{12}$$

as $c \to \infty$.

Donsker's theorem implies that

$$\sqrt{\frac{c}{r(c)}}\left(\bar{X}_{\lfloor tc/r(c)\rfloor} - z\right) \Rightarrow \sigma\,\frac{B(t)}{t}$$

as $c \to \infty$ in $D(0, \infty)$, where $B(\cdot)$ is standard Brownian motion; see [2]. A standard random time change argument (see, for example, Sect. 14 of [2]) then proves that

$$\sqrt{\frac{c}{r(c)}}\left(\bar{X}_{N(c)} - z\right) \Rightarrow \sigma\, B(1)$$

as $c \to \infty$ proving our theorem.

It remains only to prove (9). For $\gamma > -1$, write

$$\int_0^{wc_n} P(\tau_1 > y)dy = \int_{wc_n^\varepsilon}^{wc_n} P(\tau_1 > y)dy + \int_0^{wc_n^\varepsilon} P(\tau_1 > y)dy$$

for $1 > \varepsilon > 0$. On $[wc_n^\varepsilon, wc_n]$, $\log\log y/\log\log wc_n \to 1$ uniformly in y, so

$$\int_{wc_n^\varepsilon}^{wc_n} P(\tau_1 > y)dy \sim a(\log\log c_n)^\delta \int_{wc_n^\varepsilon}^{wc_n} \frac{(\log y)^\gamma}{y}dy \sim a(\log\log c_n)^\delta\,\frac{(\log c_n)^{\gamma+1}}{\gamma+1} \tag{13}$$

as $n \to \infty$. On the other hand,

$$\begin{aligned}\int_{x_0\vee 1}^{wc_n^\varepsilon} P(\tau_1 > y)dy &\le a(\log\log wc_n^\varepsilon)\int_{x_0\vee 1}^{wc_n^\varepsilon} \frac{(\log y)^\gamma}{y}dy\\ &= a(\log\log wc_n^\varepsilon)\,\frac{(\varepsilon\log c_n + \log w)^{\gamma+1}}{\gamma+1}.\end{aligned} \tag{14}$$

Since the right-hand side of (14) can be made arbitrarily small relative to the right-hand side of (13), by choosing ε small enough, we obtain (9) for $\gamma > -1$.

As for the cases where $\gamma = -1$, $\bar{F}(\cdot)$ can then be exactly integrated, and the exact integration yields the rest of (9). □

We turn next to the case where $\alpha \in (0, 1)$. To simplify our discussion, we assume here that the algorithm has been designed so that $L(x) \equiv a$ for $x \ge x_0$. For $0 < \alpha < 1$, define the constant C_α as

$$C_\alpha = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}; \tag{15}$$

here $\Gamma(\cdot)$ is the gamma function.

Theorem 2 *Suppose* $\sigma^2 = \mathrm{Var}(X_1) < \infty$ *and assume* $\alpha \in (0, 1)$. *Set* $\kappa = (a/C_\alpha)^{1/\alpha}$. *Then,*

$$(c/\kappa)^{\alpha/2}\left(\bar{X}_{N(c)} - z\right) \Rightarrow \frac{\sigma\, B(\nu_\alpha)}{\nu_\alpha}$$

as $c \to \infty$*, where* ν_α *is independent of the standard Brownian motion* B *and has the distribution of* $1/S_\alpha(1, 1, 0)^\alpha$.

Proof We start by noting that Theorem 4.5.3 of [11] implies that

$$Y_n(\cdot) \triangleq \frac{\sum_{i=1}^{\lfloor n\cdot \rfloor} \tau_i}{c_n} \Rightarrow Y_\alpha(\cdot)$$

as $n \to \infty$ in $D[0, \infty)$, where $Y_\alpha = (Y_\alpha(t) : t \geq 0)$ is a Lévy process with $Y_\alpha(1) \stackrel{\mathscr{D}}{=} S_\alpha(1, 1, 0)$ and $\stackrel{\mathscr{D}}{=}$ means "equality in distribution." The constants c_n are given by

$$c_n = (a/C_\alpha)^{1/\alpha} n^{1/\alpha} = \kappa n^{1/\alpha}.$$

Let

$$Z_n(\cdot) = \frac{\sum_{i=1}^{\lfloor n\cdot \rfloor} X_i - ze(n\cdot)}{\sqrt{n}},$$

where $e(t) = t$. We will now prove that Z_n is asymptotically independent of Y_n as $n \to \infty$. To establish the independence, we will "Poissonify." Specifically, let $R = (R(t) : t \geq 0)$ be a unit rate Poisson process with associated event times $(T_n : n \geq 1)$. Put

$$\tilde{Z}_n(t) = Z_n(R(nt)/n), \;\; \tilde{Y}_n(t) = Y_n(R(nt)/n)$$

and set

$$\begin{aligned}
\tilde{Z}_n^{(1)}(t) &= \sum_{i=1}^{R(nt)} (X_i - z) I(\tau_i \leq a_n)/n^{1/2}, \\
\tilde{Y}_n^{(1)}(t) &= \sum_{i=1}^{R(nt)} \tau_i I(\tau_i \leq a_n)/c_n, \\
\tilde{Z}_n^{(2)}(t) &= \sum_{i=1}^{R(nt)} (X_i - z) I(\tau_i > a_n)/n^{1/2}, \\
\tilde{Y}_n^{(2)}(t) &= \sum_{i=1}^{R(nt)} \tau_i I(\tau_i > a_n)/c_n.
\end{aligned}$$

Because of the Poissonification, $\tilde{Z}_n^{(1)}$ is independent of $\tilde{Y}_n^{(2)}$. Note that

$$E\left(\sup_{0\leq s\leq t} \tilde{Y}_n^{(1)}(s)\right) = E\left(\tilde{Y}_n^{(1)}(t)\right) \leq \frac{ntE(\tau_1 I(\tau_1 \leq a_n))}{c_n}.$$

If we choose $a_n = n^{1/(2\alpha)-1/2}$, we find that $nE(\tau_1 I(\tau_1 \leq a_n))/c_n \to 0$, so that

$$\sup_{0 \le s \le t} \tilde{Y}_n^{(1)}(s) \Rightarrow 0$$

as $n \to \infty$. Similarly, Kolmogorov's inequality implies that

$$\sup_{0 \le s \le t} |\tilde{Z}_n^{(2)}(s)| \Rightarrow 0$$

as $n \to \infty$. Since

$$\tilde{Z}_n = \tilde{Z}_n^{(1)} + \tilde{Z}_n^{(2)} \Rightarrow \sigma B$$

and

$$\tilde{Y}_n = \tilde{Y}_n^{(1)} + \tilde{Y}_n^{(2)} \Rightarrow Y_\alpha$$

as $n \to \infty$ in $D[0, \infty)$,

$$(\tilde{Z}_n, \tilde{Y}_n) \Rightarrow (\sigma B, Y_\alpha)$$

as $n \to \infty$, where B is independent of Y_α. We now recover Z_n and Y_n via the representation

$$Z_n(t) = \tilde{Z}_n(T_{\lfloor nt \rfloor}/n),$$
$$Y_n(t) = \tilde{Y}_n(T_{\lfloor nt \rfloor}/n).$$

Since $T_{\lfloor n \cdot \rfloor}/n \Rightarrow e(\cdot)$ in $D[0, \infty)$, it follows that

$$(Z_n, Y_n) \Rightarrow (\sigma B, Y_\alpha) \tag{16}$$

as $n \to \infty$; see Theorem 13.2.2 of [11].

If f is a bounded continuous function on $D[0, \infty)$, (16) implies that

$$E\left(f(Z_{\lfloor (c/\kappa)^\alpha \rfloor}) I(Y_{\lfloor (c/\kappa)^\alpha \rfloor}(y) > 1)\right) \to E\left(f(\sigma B) I(Y_\alpha(y) > 1)\right) \tag{17}$$

as $c \to \infty$, since $Y_\alpha(y)$ is a continuous rv (so its distribution is continuous); see [3, 11]. But

$$\{Y_{\lfloor (c/\kappa)^\alpha \rfloor}(y) > 1\} = \left\{ \sum_{i=1}^{\lfloor (c/\kappa)^\alpha y \rfloor} \tau_i > \kappa (\lfloor (c/\kappa)^\alpha \rfloor)^{1/\alpha} \right\}$$

and hence (17) implies that

$$E\left(f(Z_{\lfloor (c/\kappa)^\alpha \rfloor}) I\left(\frac{N(c)}{(c/\kappa)^\alpha} < y \right) \right) \to E(f(\sigma B)) P(Y_\alpha(y) > 1) \tag{18}$$

as $c \to \infty$. Also,

$$Y_{\lfloor (c/\kappa)^\alpha \rfloor}(y) = Y_{\lfloor (c/\kappa)^\alpha y \rfloor}(1)(y^{1/\alpha} + o(1))$$

(where $o(a_n)$ is a function for which $o(a_n)/a_n \to 0$ as $n \to \infty$), so that

$$\begin{aligned} E(f(Z_{\lfloor (c/\kappa)^\alpha y\rfloor})I(Y_{\lfloor (c/\kappa)^\alpha\rfloor}(y) > 1)) &\to E(f(\sigma B))P(Y_\alpha(1) > y^{-1/\alpha}) \\ &= E(f(\sigma B))P(S_\alpha(1,1,0) > y^{-1/\alpha}) \\ &= E(f(\sigma B))P(\nu_\alpha < y) \end{aligned} \tag{19}$$

as $c \to \infty$. Combining (18) and (19), we have that

$$E\left(f(Z_{\lfloor (c/\kappa)^\alpha y\rfloor})I\left(\frac{N(c)}{(c/\kappa)^\alpha} < y\right)\right) \to E\left(f(\sigma B)\right)P(\nu_\alpha < y)$$

as $c \to \infty$, so that

$$\left(Z_{\lfloor (c/\kappa)^\alpha\rfloor}, \frac{N(c)}{(c/\kappa)^\alpha}\right) \Rightarrow (\sigma B, \nu_\alpha)$$

as $c \to \infty$, where ν_α is independent of σB. The continuous mapping principle, based on a time substitution, then yields the theorem. □

We finish this section with a brief discussion of the rate of convergence of Monte Carlo algorithms in the setting in which $\mathrm{Var}(X_1) = \infty = E(\tau_1)$, when both X_1 and τ_1 lie in the domain of attraction of a stable law. Of course, we need $E(|X_1|) < \infty$ in order that $z = E(X_1)$ be well-defined, so we are considering here a stable index ρ for X_1 lying in the interval $(1, 2)$. To simplify our exposition, we postulate that X_1 is in the normal domain of attraction of $S_\rho(1, \beta, 0)$, so that

$$P(|X_1| > x) \sim bx^{-\rho} \tag{20}$$

as $x \to \infty$, where $b > 0$.

Let $Y_\rho = (Y_\rho(t) : t \geq 0)$ be the Lévy process in which $Y_\rho(1) \overset{\mathscr{D}}{=} S_\rho(1, \beta, 0)$.

Theorem 3 *Suppose that X_1 lies in the domain of attraction of $S_\rho(1, \beta, 0)$ and satisfies (20).*

(a) If τ_1 satisfies the hypotheses of Theorem 1, then

$$\left(\frac{c}{r(c)}\right)^{1-1/\rho}\left(\bar{X}_{N(c)} - z\right) \Rightarrow d\,Y_\rho(1)$$

as $n \to \infty$, where $d = (b/C_\rho)^{1/\rho}$.

(b) If τ_1 satisfies the hypotheses of Theorem 2 (so that $\alpha \in (0, 1)$), then

$$\left(\frac{c}{\kappa}\right)^{\alpha(1-1/\rho)}\left(\bar{X}_{N(c)} - z\right) \Rightarrow \frac{d\,Y_\rho(\nu_\alpha)}{\nu_\alpha}$$

as $c \to \infty$, where Y_ρ is independent of ν_α.

Proof We note that under our hypotheses,

$$\left(\frac{c}{r(c)}\right)^{1-1/\rho}\left(\bar{X}_{\lfloor \frac{c}{r(c)}\cdot \rfloor}-z\right)\Rightarrow \frac{d\,Y_\alpha(\cdot)}{e(\cdot)}$$

in $D[0,\infty)$. We now utilize (12) and the stochastic continuity of Y_ρ to apply the continuous mapping principle, thereby obtaining (a).

For part (b), we argue as in Theorem 2 that

$$\left(n^{1-1/\rho}\left(\bar{X}_{\lfloor n\cdot\rfloor}-z\right),\ \frac{\sum_{i=1}^{\lfloor n\cdot\rfloor}\tau_i}{n^{1/\alpha}}\right)\Rightarrow\left(\frac{d\,Y_\rho(\cdot)}{e(\cdot)},\kappa Y_\alpha(\cdot)\right)$$

as $n\to\infty$, where Y_ρ and Y_α are independent and Y_α is as in Theorem 2. It follows that

$$\left((c/\kappa)^{\alpha(1-1/\rho)}\left(\bar{X}_{\lfloor (c/\kappa)^\alpha\cdot\rfloor}-z\right),\ \frac{N(c)}{(c/\kappa)^\alpha}\right)\Rightarrow\left(\frac{d\,Y_\rho(\cdot)}{e(\cdot)},\nu_\alpha\right)$$

as $c\to\infty$, where ν_α is independent of Y_ρ. We finish the proof with a continuous mapping argument based on use of the obvious composition mapping. □

We remark that this theorem is more challenging to apply in the Monte Carlo setting, than are Theorems 1 and 2, because it requires verifying that X_1 is in the domain of attraction of a stable law.

3 Applications to Debiased MLMC

Suppose that $z=E(W)$, where W can not be simulated in finite time, but an approximating sequence $(W_n : n\geq 1)$ is available, in which the W_n's can be simulated in finite time. In particular, suppose that W_n converges to W in L^2, so that

$$\|W_n-W\|_2\to 0$$

as $n\to\infty$, where $\|U\|_2=\sqrt{E(U^2)}$ for a generic rv U.

Set $W_0=0$ and put $\Delta_i=W_i-W_{i-1}$ for $i\geq 1$. Then, under appropriate regularity conditions (see below),

$$X=\sum_{i=1}^{M}\frac{\Delta_i}{P(M\geq i)} \tag{21}$$

is an unbiased estimator for z, when M is generated independently of the Δ_i's. Specifically, Theorem 1 of [10] shows that if

$$\sum_{n=1}^{\infty} \frac{\|W_{n-1} - W\|_2^2}{P(M \geq n)} < \infty,$$

then X is unbiased, and

$$E(X^2) = \sum_{n=1}^{\infty} \frac{\|W_{n-1} - W\|_2^2 - \|W_n - W\|_2^2}{P(M \geq n)}.$$

An important application of such "debiased MLMC" estimator is numerical computation for stochastic differential equations (SDE's). In that context, the simplest and most natural approximation to W is to use the sequence $\{W_n : n \geq 1\}$ obtained by Euler discretization of the SDE. Specifically, we let W_n be the Euler discretization to W associated with a time step of order 2^{-n}, and couple the W_n's via the use of a common driving Brownian motion across all the approximations in n. If we do this, it is known in significant generality that for problems involving Lipschitz functions of the final value, $\|W_n - W\|_2^2 = O(2^{-n})$ as $n \to \infty$; see [8].

Hence, a sufficient condition on the distribution of M ensuring that $\mathrm{Var}(X) < \infty$ is to choose M so that

$$\sum_{n=1}^{\infty} \frac{2^{-n}}{P(M \geq n)} < \infty. \tag{22}$$

However, we also need to consider the computer time τ for generating X. If we take the (reasonable) view that generating a discretization with time step 2^{-n} takes computational effort 2^n, then $\tau = 2^M$. So,

$$P(\tau > x) = P(M > \lfloor \log_2 x \rfloor).$$

Suppose we now choose M so that $P(M > n) = 2^{-\alpha n}$ for $n \geq n_0$, with $\alpha \in (0, 1)$; this choice of α guarantees that $\mathrm{Var}(X) < \infty$. Hence, $P(\tau > x) = 2^{-\lfloor \log_2 x \rfloor \alpha}$ for x sufficiently large. However, τ does not have a regular varying tail, so the theory of Sect. 2 does not directly apply. But we can always choose to randomly delay the completion time of X. Specifically suppose that we start by generating τ so that

$$P(\tau > x) = x^{-\alpha} \tag{23}$$

for $x \geq 1$. With τ in hand, we set $M = \lfloor \log_2 \tau \rfloor$. Note that $P(M > i) = P(\tau \geq 2^i)$ so $P(M \geq i) = P(\tau \geq 2^{i-1}) = 2^{(1-i)\alpha}$ for $i \geq 1$. We now delay the completion of X from time $2^M = 2^{\lfloor \log_2 \tau \rfloor}$ to time τ. With this convention in place, our theory applies and Theorem 2 establishes that

$$(c/\kappa)^{\alpha/2} \left(\bar{X}_{N(c)} - z\right) \Rightarrow \frac{\sigma\, B(\nu_\alpha)}{\nu_\alpha}$$

as $c \to \infty$, where $\kappa = C_\alpha^{-1/\alpha}$. Hence, the rate of convergence of $\bar{X}_{N(c)}$ to z is of order $c^{-\alpha/2}$ with this choice of randomization for M.

The above CLT-type theorem also allows us to construct confidence intervals for z in this setting in which $E(\tau) = \infty$. In particular, if we select $\tilde{a}$ such that $P(-\tilde{a} \le B(\nu_\alpha)/\nu_\alpha \le \tilde{a}) = 0.9$ (say), then the interval

$$\left[\bar{X}_{N(c)} - \tilde{a}\hat{\sigma}(c)\left(\frac{\kappa}{c}\right)^{\alpha/2},\ \bar{X}_{N(c)} + \tilde{a}\hat{\sigma}(c)\left(\frac{\kappa}{c}\right)^{\alpha/2}\right] \tag{24}$$

is an asymptotic 90% confidence interval for z, where $\hat{\sigma}(c)$ is the sample standard deviation estimator given by

$$\hat{\sigma}(c) = \sqrt{\frac{1}{N(c)-1}\sum_{i=1}^{N(c)}\left(X_i - \bar{X}_{N(c)}\right)^2}. \tag{25}$$

Other choices for the randomization distribution are also possible. In the case α, suppose that we generate τ so that

$$P(\tau > x) = x^{-1}(\log x)^\gamma(\log\log x)^\delta$$

for x sufficiently large. Again, we let $M = \lfloor \log_2 \tau \rfloor$ and again note that $P(M > i) = P(\tau \ge 2^{i-1})$ for $i \ge 1$. In order that $\mathrm{Var}(X) < \infty$, we choose either $\gamma > 1$ or $\gamma = 1$ with $\delta > 1$. Applying Theorem 1, we find that

$$\sqrt{\frac{c}{r(c)}}\left(\bar{X}_{N(c)} - z\right) \Rightarrow \sigma\, N(0,1) \tag{26}$$

as $c \to \infty$, where $r(c) = (1+\gamma)^{-1}(\log c)^{1+\gamma}(\log\log c)^\delta$. The best convergence rate is attained when $\gamma = 1$ with $\delta > 1$ but close to 1. In this case, the exact convergence rate is of order $c^{-1/2}(\log c)(\log\log c)^{\delta/2}$, and the computational budget required to obtain an accuracy ε is of order $\varepsilon^{-2}(\log(1/\varepsilon))^2(\log\log(1/\varepsilon))^\delta$ with $\delta > 1$. This complexity estimate for debiased MLMC is slightly better than that provided by Proposition 4 of [10], in which the estimate takes the form $\varepsilon^{-2}(\log(1/\varepsilon))^q$ with $q > 2$.

As for the case where τ is chosen so that (23) holds, confidence intervals for z can again be generated. The CLT (26) implies that if $\tilde{a}$ is chosen so that $P(-\tilde{a} \le N(0,1) \le \tilde{a}) = 0.9$, then

$$\left[\bar{X}_{N(c)} - \tilde{a}\hat{\sigma}(c)\sqrt{\frac{r(c)}{c}},\ \bar{X}_{N(c)} + \tilde{a}\hat{\sigma}(c)\sqrt{\frac{r(c)}{c}}\right]$$

is an asymptotic 90% confidence interval for z, as $c \to \infty$.

If one prefers an analysis in which no delay in generating X is introduced, one can observe that $\tau/2 \le 2^M \le \tau$ when $M = \lfloor \log_2 \tau \rfloor$. If $\underset{\sim}{N}(c) = \max\{n \ge 0 : \sum_{i=1}^n \tau_i \le c\}$ and we model the time required to generate X_i as 2^{M_i}, then $\underset{\sim}{N}(c) \le N(c) \le \underset{\sim}{N}(2c)$ for $c \ge 0$. Furthermore, when $\alpha = 1$ so that Theorem 1 applies, then

$$\sqrt{\frac{c}{r(c)}}|\bar{X}_{N(c)} - z| \le \sqrt{\frac{c}{r(c)}}|\bar{X}_{\underset{\sim}{N}(c)} - z|\frac{\underset{\sim}{N}(2c)}{\underset{\sim}{N}(c)} + \sqrt{\frac{c}{r(c)}}\frac{|\sum_{k=\underset{\sim}{N}(c)}^{N(c)}(X_i - z)|}{N(c)}. \tag{27}$$

Now, Theorem 1 applies to $(c/r(c))^{1/2}|\bar{X}_{\underset{\sim}{N}(c)} - z|$ and so is stochastically bounded (i.e. tight) in c. In addition, the proof of Theorem 1 shows that both $\underset{\sim}{N}(c)r(c)/c$ and $\underset{\sim}{N}(2c)r(c)/c$ are tight, so that the first term on the right-hand side of (27) is stochastically bounded. Furthermore, Kolmogorov's maximal inequality and $\mathrm{Var}(X) < \infty$ imply that

$$\sqrt{\frac{r(c)}{c}}|\sum_{k=\underset{\sim}{N}(c)}^{N(c)}(X_i - z)| \le \sqrt{\frac{r(c)}{c}} \max_{\underset{\sim}{N}(c) \le k \le \underset{\sim}{N}(2c)} |\sum_{i=\underset{\sim}{N}(c)}^{k}(X_i - z)|$$

is stochastically bounded, so that the tightness of $N(c)r(c)/c$ yields the stochastic boundedness of the left-hand side of (27).

This proves that $\bar{X}_{N(c)}$ (with no delay introduced) does indeed converge to z at a rate that is at most $(r(c)/c)^{1/2}$ as $c \to \infty$. Note, however, that we can not get an asymptotic confidence interval for z directly from this bounding argument.

4 A Numerical Example

In this section, we implement a debiased MLMC estimator with finite variance and infinite expected computer time and use our theory to construct asymptotically valid confidence intervals. We consider an option pricing problem in the SDE context, where the underlying diffusion process obeys the SDE

$$dX(t) = rX(t)dt + \sigma X(t)dB(t),$$

in which the parameters are the interest rate $r = 0.05$, volatity $\sigma = 0.2$ and initial asset price $X(0) = 100$. We focus on pricing a European call option with payoff $\max(X(t) - K, 0)$ at maturity $t = 1$ at three different strike prices $K = 90, 100, 110$. We implement the debiased MLMC estimator described in Sect. 3 Eq. (21), in which the approximating sequence $(W_n : n \ge 1)$ is obtained by Euler discretization with step size $2^{-n}t$ and the integer-valued randomization M is chosen as $P(M > n) = 2^{-2\alpha n}$ for $n \ge 1$ with $\alpha = 1/2$. We delay the completion time such that it has a regular

Table 1 Computational result for a debiased MLMC estimator with $E(\tau) = \infty$ and $\mathrm{Var}(X) < \infty$

Strike price	True value	Computation budget	Debiased MLMC estimator	
			C.I.	Coverage (%)
$K = 90$	16.70	20,000	16.75 ± 4.46	89.0 ± 1.62
		80,000	16.73 ± 3.21	89.4 ± 1.60
		3,20,000	16.71 ± 2.29	90.1 ± 1.55
$K = 100$	10.45	20,000	10.41 ± 3.72	88.9 ± 1.63
		80,000	10.43 ± 2.70	88.8 ± 1.67
		3,20,000	10.43 ± 1.83	89.3 ± 1.60
$K = 110$	6.04	20,000	6.01 ± 2.95	87.9 ± 1.69
		80,000	6.09 ± 2.13	88.9 ± 1.63
		3,20,000	6.05 ± 1.52	91.2 ± 1.47

varying tail $P(\tau > x) = x^{-\alpha}$. Our theory applies and Theorem 2 shows that

$$(c/\kappa)^{1/4} \left(\bar{X}_{N(c)} - z\right) \Rightarrow \frac{\sigma\, B(\nu_\alpha)}{\nu_\alpha}$$

as $c \to \infty$, where $\kappa = C_\alpha^{-1/\alpha}$ and C_α is defined in Eq. (15). This result establishes an exact convergence rate of order $c^{-1/4}$ for the estimator and allows us to construct confidence intervals following the procedure in Sect. 3; see Eqs. (24) and (25). For each strike price, we implement the algorithm with computational budget $c = 10000$, 20000, 80000 and 320000. Finally, in each experiment, we construct an approximate 90% confidence interval for the mean, based on the limit distribution above, and then run 1000 independent replications of each experiment.

In Table 1, we report the computational results. The columns labeled C.I. report the average midpoint of the 1000 intervals, together with the average confidence interval half-width, again averaged over the 1000 replications. The columns labeled Coverage report 90% confidence intervals (based on the normal approximation) for the percentage of the 1000 replications in which the confidence interval contains the true option price. As shown in the table, the confidence intervals are asymptotically valid. We further note that this debiased MLMC estimator (with parameter $\alpha = 1/2$) demonstrates a convergence rate of order $c^{-1/4}$, as the length of the confidence interval roughly halves when the sample size is multiplied by a factor of sixteen. This result agrees with the exact convergence rate established by our CLT.

Acknowledgements The authors gratefully acknowledge the support of the National Science Foundation under Award Number DMS-1320158.

References

1. Alaya, M.B., Kebaier, A.: Central limit theorem for the multilevel Monte Carlo Euler method. Ann. Appl. Probab. **25**(1), 211–234 (2015)
2. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
3. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation, vol. 27. Cambridge University Press, Cambridge (1989)
4. Glynn, P.W., Heidelberger, P.: Bias properties of budget constrained simulations. Oper. Res. **38**(5), 801–814 (1990)
5. Glynn, P.W., Whitt, W.: The asymptotic efficiency of simulation estimators. Oper. Res. **40**(3), 505–520 (1992)
6. Giorgi, D., Lemaire, V., Pagès, G.: Limit theorems for weighted and regular multilevel estimators. Monte Carlo Methods Appl. **23**(1), 43–70 (2017)
7. Hammersley, J., Handscomb, D.: Monte Carlo Methods. Methuen, London (1964)
8. Kloeden, P.E., Planten, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)
9. Lemaire, V., Pagès, G.: Multilevel Richardson-Romberg extrapolation. Bernoulli **23**(4A), 2643–2692 (2017)
10. Rhee, C.-H., Glynn, P.W.: Unbiased estimation with square root convergence for SDE models. Oper. Res. **63**(5), 1026–1043 (2015)
11. Whitt, W.: Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues. Springer Science & Business Media, New York (2002)