ESAIM: PROCEEDINGS AND SURVEYS, September 2017, Vol. 59, p. 104-114
B. Bouchard, E. Gobet and B. Jourdain, Editors

A CLT FOR INFINITELY STRATIFIED ESTIMATORS, WITH APPLICATIONS TO DEBIASED MLMC *

Zeyu Zheng[1] and Peter W. Glynn[2]

Abstract. This paper develops a general central limit theorem (CLT) for post-stratified Monte Carlo estimators with an associated infinite number of strata. In addition, consistency of the corresponding variance estimator is established in the same setting. With these results in hand, one can then construct asymptotically valid confidence interval procedures for such infinitely stratified estimators. We then illustrate our general theory, by applying it to the specific case of debiased multi-level Monte Carlo (MLMC) algorithms. This leads to the first asymptotically valid confidence interval procedure for such stratified debiased MLMC procedures.

1. Introduction

Suppose that our goal is to compute $\alpha = EX$, and that the sample space Ω underlying X can be partitioned into events $A_1, A_2, \ldots$ for which $p_i \triangleq P(A_i)$ is known for $i \geq 1$. Let $I_i = I(A_i)$ be the indicator random variable (rv) corresponding to A_i, and suppose that $(X_1, I_{i1} : i \geq 1)$, $(X_2, I_{i2} : i \geq 1), \ldots$ is an independent and identically distributed (iid) sequence of copies of $(X, I_i : i \geq 1)$. The conventional (crude) Monte Carlo estimator for α is then

$$\bar{X}_n \triangleq \frac{1}{n}\sum_{j=1}^{n} X_j = \sum_{i=1}^{\infty} I(N_n(i) \geq 1) \cdot \frac{N_n(i)}{n} \cdot \frac{\sum_{j=1}^{n} X_j I_{ij}}{N_n(i)}, \tag{1}$$

where $N_n(i) \triangleq \sum_{j=1}^{n} I_{ij}$ is the number of times the event A_i occurs in the first n samples generated.

The idea behind *stratification* is to exploit our knowledge of the p_i's. In particular, the *post-stratified* version of $\bar{X}_n$ replaces $N_n(i)/n$ by p_i, thereby yielding the estimator

$$\mathcal{P}_n \triangleq \sum_{i=1}^{\infty} p_i \cdot \frac{\sum_{j=1}^{n} X_j I_{ij}}{N_n(i)} \cdot I(N_n(i) \geq 1); \tag{2}$$

the events $A_1, A_2, \ldots$ are then known as the *strata* associated with the estimator $\mathcal{P}_n$. When the number of strata is finite, it is known that $\mathcal{P}_n$ exhibits smaller variance than does $\bar{X}_n$; see p.151 of [Asmussen and Glynn(2007)].

In this paper, we provide a central limit theorem (CLT) for $\mathcal{P}_n$. The novelty here is an extension of the CLT to the case of infinitely many strata; CLT's for the case in which the number of strata is finite (i.e., $A_1, A_2, \ldots, A_m$ partition Ω for $m < \infty$) are known; see, for example, p.151 of [Asmussen and Glynn(2007)].

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

[1] Department of Management Science and Engineering, Stanford University, Stanford, California 94305, zyzheng@stanford.edu
[2] Department of Management Science and Engineering, Stanford University, Stanford, California 94305, glynn@stanford.edu

The technical complication in the infinite case involves controlling the infinitely many random ratio estimators implicit in (2). This CLT allows us to construct asymptotically valid fixed sample size confidence intervals for α.

Our main motivation for proving such CLT's for infinitely stratified estimators has to do with their application to debiased multi-level Monte Carlo (MLMC) estimators. This class of algorithms was first introduced by [Rhee and Glynn(2015)] and involves a randomization that produces an unbiased estimator for α; see also [McLeish(2011)] and [Rhee and Glynn(2012)]. It has recently been noted by [Vihola(2015)] that the variance of these debiased MLMC estimators can be reduced by taking advantage of stratification. However, in this setting, the estimator is invariably infinitely stratified, and no corresponding CLT is provided there. This paper fills in this missing gap. Such a CLT plays an essential role in constructing asymptotically valid confidence intervals for α, so we also provide here the first asymptotically valid confidence interval procedure for stratified debiased MLMC.

This paper is organized as follows. Section 2 provides the necessary CLT for infinitely stratified estimators, while Section 3 provides conditions under which the associated variance estimator converges consistently. Section 4 concludes the paper with a brief discussion of the application to debiased MLMC.

2. The CLT for Post-stratified Estimators with Infinitely Many Strata

Assume that $E|X| < \infty$, set $\mu_i = E(X|A_i)$ and $\sigma_i^2 = \mathrm{Var}(X|A_i)$ for $i \geq 1$, and put $V_j = \sum_{i=1}^{\infty}(X_j - \mu_i)I_{ij}$ for $j \geq 1$. Also, let $\mathcal{G}_n = \sigma(N_n(i) : i \geq 1)$ be the σ-algebra generated by the $N_n(i)$'s. Then,

$$E[\mathcal{P}_n|\mathcal{G}_n] = \sum_{i=1}^{\infty} p_i \mu_i I(N_n(i) \geq 1),$$

so that

$$\mathcal{P}_n - \alpha = \mathcal{P}_n - E[\mathcal{P}_n|\mathcal{G}_n] - \sum_{i=1}^{\infty} p_i \mu_i I(N_n(i) = 0). \tag{3}$$

Observe that

$$\begin{aligned}
\mathcal{P}_n - E[\mathcal{P}_n|\mathcal{G}_n] &= \sum_{i=1}^{\infty} p_i \left(\frac{\sum_{j=1}^{n} X_j I_{ij}}{N_n(i)} - \mu_i \right) I(N_n(i) \geq 1)) \\
&= \sum_{i=1}^{\infty} p_i \left(\frac{\sum_{j=1}^{n} (X_j - \mu_i) I_{ij}}{N_n(i)} \right) I(N_n(i) \geq 1)) \\
&\overset{D}{\approx} \frac{1}{n} \sum_{i=1}^{\infty} \sum_{j=1}^{n} (X_j - \mu_i) I_{ij} \\
&= \frac{1}{n} \sum_{j=1}^{n} V_j,
\end{aligned}$$

where $\overset{D}{\approx}$ denotes "has approximately the same distribution as" (and is intended to carry no rigorous meaning). This suggests that if $\sigma^2 \triangleq \mathrm{Var}\ V_1 = \sum_{i=1}^{\infty} p_i \sigma_i^2 < \infty$, the CLT (and the observation that $EV_1 = 0$) should imply that

$$\sqrt{n}(\mathcal{P}_n - E[\mathcal{P}_n|\mathcal{G}_n]) \Rightarrow \sigma N(0,1) \tag{4}$$

as $n \to \infty$, where $\Rightarrow$ denotes weak convergence and $N(0,1)$ is a mean zero normal rv with variance 1.

Our first theorem makes (4) rigorous.

THEOREM 1. *If $\sum_{i=1}^{\infty} p_i \sigma_i^2 < \infty$ and $E|X| < \infty$, then (4) holds.*

Before proceeding to the proof, for $i \geq 1$, let $(X_j(i) : j \geq 1)$ be an independent sequence of iid rv's with common distribution $P(X \in \cdot | A_i)$, and suppose that $(X_j(i) : i, j \geq 1)$ is also independent of $(N_n(i) : i, n \geq 1)$. Note that $(\mathcal{P}_n : n \geq 1) \overset{D}{=} (\mathcal{P}'_n : n \geq 1)$, where

$$\mathcal{P}'_n = \sum_{i=1}^{\infty} p_i \frac{\sum_{j=1}^{N_n(i)} X_j(i)}{N_n(i)} I(N_n(i) \geq 1)$$

and $\overset{D}{=}$ denotes "equality in distribution." So, we can prove Theorem 1 with $\mathcal{P}'_n$ replacing $\mathcal{P}_n$.

Proof. Set $\tilde{X}_j(i) = X_j(i) - \mu_i$ for $i, j \geq 1$ and put $\tilde{\mathcal{P}}'_n = \mathcal{P}'_n - E[\mathcal{P}_n | \mathcal{G}_n] = \mathcal{P}'_n - E[\mathcal{P}'_n | \mathcal{G}_n]$. Then,

$$\tilde{\mathcal{P}}'_n = \beta_n + \frac{1}{n} \sum_{i=1}^{\infty} \sum_{j=1}^{N_n(i)} \tilde{X}_j(i)$$

where

$$\beta_n = \sum_{i=1}^{\infty} \left(\frac{p_i}{N_n(i)} - \frac{1}{n} \right) \sum_{j=1}^{N_n(i)} \tilde{X}_j(i) I(N_n(i) \geq 1).$$

Note that

$$\frac{1}{n} \sum_{i=1}^{\infty} \sum_{j=1}^{N_n(i)} \tilde{X}_j(i) \overset{D}{=} \frac{1}{n} \sum_{j=1}^{n} V_j,$$

so (4) follows if we can show that $n^{1/2} \beta_n \Rightarrow 0$ as $n \to \infty$.

For $m \geq 1$, set

$$\beta_{nm} = \sum_{i=1}^{m} p_i \frac{\sum_{j=1}^{N_n(i)} \tilde{X}_j(i)}{N_n(i)} I(N_n(i) \geq 1) - \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{N_n(i)} \tilde{X}_j(i)$$

and note that

$$\sum_{i=1}^{m} p_i \frac{\sum_{j=1}^{N_n(i)} |\tilde{X}_j(i)|}{N_n(i)} I(N_n(i) \geq 1) + \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{N_n(i)} |\tilde{X}_j(i)|$$

is square-integrable (since $N_n(i) \leq n$ and $\sigma_i^2 < \infty$ for $i \geq 1$). It follows that

$$E\beta_{nm}^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} E \left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n} \right) \left(\frac{p_j}{N_n(j)} - \frac{1}{n} \right) \sum_{k=1}^{N_n(i)} \tilde{X}_k(i) \sum_{l=1}^{N_n(j)} \tilde{X}_l(j) \, I(N_n(i) \geq 1, N_n(j) \geq 1) \right].$$

But

$$\begin{aligned}
&E \left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n} \right) \left(\frac{p_j}{N_n(j)} - \frac{1}{n} \right) \sum_{k=1}^{N_n(i)} \tilde{X}_k(i) \sum_{l=1}^{N_n(j)} \tilde{X}_l(j) \, I(N_n(i) \geq 1, N_n(j) \geq 1) | \mathcal{G}_n \right] \\
&= \left(\frac{p_i}{N_n(i)} - \frac{1}{n} \right) \left(\frac{p_j}{N_n(j)} - \frac{1}{n} \right) I(N_n(i) \geq 1, N_n(j) \geq 1) E \left[\sum_{k=1}^{N_n(i)} \tilde{X}_k(i) \sum_{l=1}^{N_n(j)} \tilde{X}_l(j) | \mathcal{G}_n \right] \\
&= 0
\end{aligned}$$

for $i \neq j$, since the $\tilde{X}_k(i)$'s are independent of the $\tilde{X}_l(j)$'s for $i \neq j$ and $E\tilde{X}_k(i) = 0$ for $i, k \geq 1$. So,

$$E\beta_{nm}^2 = \sum_{i=1}^{m} E\left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 \left(\sum_{k=1}^{N_n(i)} \tilde{X}_k(i)\right)^2 I(N_n(i) \geq 1)\right].$$

Furthermore,

$$\beta_{nm}^2 \to \beta_n^2 \quad a.s.$$

as $m \to \infty$, so Fatou's lemma implies that

$$\begin{aligned} E\beta_n^2 &\leq \varliminf_{m\to\infty} E\beta_{nm}^2 \\ &\leq \varliminf_{m\to\infty} \sum_{i=1}^{m} E\left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 \left(\sum_{k=1}^{N_n(i)} \tilde{X}_k(i)\right)^2 I(N_n(i) \geq 1)\right] \\ &= \sum_{i=1}^{\infty} E\left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 \left(\sum_{k=1}^{N_n(i)} \tilde{X}_k(i)\right)^2 I(N_n(i) \geq 1)\right], \end{aligned} \tag{5}$$

where we used the fact that the finite sum (to m) increases monotonically to the infinite sum for the final step.

Note that

$$\begin{aligned} &E\left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 \left(\sum_{k=1}^{N_n(i)} \tilde{X}_k(i)\right)^2 I(N_n(i) \geq 1) \,\middle|\, \mathcal{G}_n\right] \\ &= \left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 N_n(i)\sigma_i^2 I(N_n(i) \geq 1) \\ &= \frac{(p_i n - N_n(i))^2}{N_n(i) n^2} \sigma_i^2 I(N_n(i) \geq 1) \\ &\leq \frac{(N_n(i) - p_i n)^2}{n^2} \sigma_i^2. \end{aligned}$$

Consequently,

$$\begin{aligned} &n\, E\left[\left(\frac{p_i}{N_n(i)} - \frac{1}{n}\right)^2 \left(\sum_{k=1}^{N_n(i)} \tilde{X}_k(i)\right)^2 I(N_n(i) \geq 1)\right] \\ &\leq \frac{1}{n} E(N_n(i) - p_i n)^2 \sigma_i^2 = p_i(1 - p_i)\sigma_i^2 \leq p_i \sigma_i^2. \end{aligned} \tag{6}$$

Also,

$$
\begin{aligned}
& n\,E\left[\left(\frac{p_i}{N_n(i)}-\frac{1}{n}\right)^2\left(\sum_{k=1}^{N_n(i)}\tilde{X}_k(i)\right)^2 I(N_n(i)\geq 1)\right] \\
& = \sigma_i^2 E\left[\frac{(p_i n - N_n(i))^2}{N_n(i)n} I(N_n(i)\geq 1)\right] \\
& = \sigma_i^2 E\left[\frac{(p_i n - N_n(i))^2}{N_n(i)n} I(N_n(i)\geq p_i n/2)\right] + \sigma_i^2 E\left[\frac{(p_i n - N_n(i))^2}{N_n(i)n} I(1\leq N_n(i) < p_i n/2)\right] \\
& \leq 2\sigma_i^2 \frac{E\,(p_i n - N_n(i))^2}{n^2 p_i} + \sigma_i^2 n\,E\left[\frac{1}{N_n(i)} I(1\leq N_n(i) < p_i n/2)\right] \\
& \leq \frac{2\sigma_i^2}{n} + \sigma_i^2 n\,P(1\leq N_n(i) < p_i n/2) \to 0
\end{aligned} \tag{7}
$$

as $n\to\infty$, where we apply a standard large deviations bound at the final step to conclude that $P(N_n(i)\leq p_i n/2)$ converges to 0 exponentially rapidly in n; see, for example, p.95 of [Serfling(1980)].

Since $\sum_{i=1}^{\infty} p_i\sigma_i^2 < \infty$, (6) and (7) imply that we may apply the Dominated Convergence Theorem to obtain the conclusion

$$
n\sum_{i=1}^{\infty} E\left[\left(\frac{p_i}{N_n(i)}-\frac{1}{n}\right)^2\left(\sum_{k=1}^{N_n(i)}\tilde{X}_k(i)\right)^2 I(N_n(i)\geq 1)\right]\to 0
$$

as $n\to\infty$. The inequality (5) then establishes that $n\,E\beta_n^2\to 0$ as $n\to\infty$. Finally, Markov's inequality shows that $n^{1/2}\beta_n \Rightarrow 0$ as $n\to\infty$, proving the theorem. □

We note that Theorem 1's centering is not $\alpha = EX$, but instead involves the random centering $E[\mathcal{P}_n|\mathcal{G}_n]$. Since our main rationale for proving a CLT is its application to construction of approximate confidence intervals for α, we need to now obtain conditions under which the random centering can be replaced by α. This requires proving that the remainder term in (3) converges to zero. In other words, we now need to study conditions under which

$$
n^{1/2}\sum_{i=1}^{\infty} p_i\mu_i I(N_n(i)=0)\Rightarrow 0 \tag{8}
$$

as $n\to\infty$.

A natural additional condition to consider is $\sum_{i=1}^{\infty} p_i\mu_i^2 < \infty$, in view of the identity

$$
\begin{aligned}
\mathrm{Var}(X) &= E\left[\mathrm{Var}(X|I_j : j\geq 1)\right] + \mathrm{Var}\left(E[X|I_j : j\geq 1]\right) \\
&= \sum_{i=1}^{\infty} p_i\sigma_i^2 + \left(\sum_{i=1}^{\infty} p_i\mu_i^2 - \alpha^2\right);
\end{aligned}
$$

see, for example, p.56 of [Bratley, Fox and Schrage(1987)]. However, this condition fails to be sufficient to guarantee (8).

Example 1. We start by observing that

$$\begin{aligned}
& n \text{ Var}\left(\sum_{i=1}^{\infty} p_i\mu_i I(N_n(i)=0)\right) \\
&= n\sum_{i=1}^{\infty} p_i^2\mu_i^2\left[(1-p_i)^n-(1-p_i)^{2n}\right] + 2n\sum_{i<j} p_ip_j\mu_i\mu_j(1-p_i-p_j)^n\left[1-\left(\frac{1-p_i-p_j+p_ip_j}{1-p_i-p_j}\right)^n\right] \\
&\leq n\sum_{i=1}^{\infty} p_i^2\mu_i^2(1-p_i)^n = n\sum_{i=1}^{\infty} p_i^2\mu_i^2\exp(n\log(1-p_i)) \\
&\leq \sum_{i=1}^{\infty} p_i\mu_i^2(np_i)e^{-np_i},
\end{aligned}$$

But $xe^{-x} \leq e^{-1}$ for $x \geq 0$, so the i-th term is bounded by $p_i\mu_i^2e^{-1}$, uniformly in n. On the other hand, $np_i^2\mu_i^2e^{-np_i} \to 0$ as $n \to \infty$, for each $i \geq 1$. So, if $\sum_{i=1}^{\infty} p_i\mu_i^2 < \infty$, the Dominated Convergence Theorem applies, and

$$n \text{ Var}\left(\sum_{i=1}^{\infty} p_i\mu_i I(N_n(i)=0)\right) \to 0 \tag{9}$$

as $n \to \infty$.

We now consider the case in which $p_i = ci^{-p}$ for $c > 0$ and $p > 1$, with $\mu_i = i^q$. We require that $2q < p-1$, in order that $\sum_{i=1}^{\infty} p_i\mu_i^2 < \infty$. Note that

$$\begin{aligned}
& n^{1/2}E\sum_{i=1}^{\infty} p_i\mu_i I(N_n(i)=0) \\
&= cn^{1/2}\sum_{i=1}^{\infty} i^{-p+q}\exp(n\log(1-p_i)) \\
&\geq cn^{1/2}\sum_{i:p_i\leq\frac{1}{2n}} i^{-p+q}\exp(-2np_i) \\
&\geq ce^{-1}n^{1/2}\sum_{i:p_i\leq\frac{1}{2n}} i^{-p+q} \\
&= \Omega(n^{\frac{1}{2}+\frac{1}{p}(1-p+q)}) = \Omega(n^{\frac{1}{p}+\frac{q}{p}-\frac{1}{2}}),
\end{aligned}$$

where $\Omega(a_n)$ denotes a sequence which is bounded below by a positive multiple of $(a_n : n \geq 1)$. It follows that if $(1+q)/p \geq 1/2$ (e.g. $q = 0$ and $1 < p \leq 2$), $\underline{\lim}_{n\to\infty} n^{1/2}E\sum_{i=1}^{\infty} p_i\mu_i I(N_n(i)=0)$ is positive. In view of (9) and Chebyshev's inequality, we conclude that (8) does not hold for such examples.

The above example makes clear that Var $X < \infty$ does not guarantee that the CLT for our post-stratified estimator holds with α as its centering. Different conditions are required.

Let J be a rv with probability mass function given by $\{p_i : i \geq 1\}$.

PROPOSITION 1. *Suppose that* $E|\mu_J|^r < \infty$ *and* $EJ^\alpha < \infty$, *where* $\alpha > 1$ *and* $r > \frac{2\alpha}{\alpha-1}$. *Then*

$$n^{1/2}\sum_{i=1}^{\infty} \mu_i p_i I(N_n(i)=0) \Rightarrow 0$$

as $n \to \infty$.

Proof. We start by applying Hölder's inequality, so that

$$n^{1/2}\sum_{i=1}^{\infty}|\mu_i|p_i I(N_n(i)=0) \le \left(\sum_{i=1}^{\infty}|\mu_i|^r p_i\right)^{\frac{1}{r}} \cdot \left(n^{\frac{r}{2(r-1)}}\sum_{i=1}^{\infty}p_i I(N_n(i)=0)\right)^{\frac{r-1}{r}}.$$

Then, the result follows from Markov's inequality if we prove that

$$n^{\frac{r}{2(r-1)}}\sum_{i=1}^{\infty}p_i P(N_n(i)=0) = n^{\frac{r}{2(r-1)}}\sum_{i=1}^{\infty}p_i(1-p_i)^n \to 0 \tag{10}$$

as $n \to \infty$.

Now, Markov's inequality and the fact that $xe^{-x} \le e^{-1}$ for $x \ge 0$ imply that for $a > 0$,

$$\begin{aligned}\sum_{i=1}^{\infty}p_i(1-p_i)^n &\le \sum_{i=1}^{\infty}p_i\exp(-np_i)\\ &\le n^{-1}\sum_{i\le n^a}(np_i)\exp(-np_i) + P(J>n^a)\\ &\le n^{-1+a}e^{-1} + n^{-a\alpha}EJ^{\alpha},\end{aligned}$$

so (10) follows if we can find $a > 0$ such that $r(2(r-1))^{-1} - 1 + a < 0$ and $r(2(r-1))^{-1} - a\alpha < 0$. This requires that $r(2(r-1)\alpha)^{-1} < a < 1 - r(2(r-1))^{-1}$. The inequality $r(2(r-1)\alpha)^{-1} < 1 - r(2(r-1))^{-1}$ is equivalent to requiring $r > 2\alpha/(\alpha-1)$, proving the result. □

Note that if the μ_i's are bounded, all that is needed is $\alpha > 1$ (since we can choose r arbitrarily large in the bounded case). This is equivalent to requiring that $p > 2$ in the context of Example 1, so that Proposition 1's conditions are sharp in that case.

Theorem 1 and Proposition 1 together provide our desired CLT for $\mathcal{P}_n$ with centering given by α.

THEOREM 2. *If $\sum_{i=1}^{\infty}p_i\sigma_i^2 < \infty$, $E|\mu_J|^r < \infty$, and $EJ^{\alpha} < \infty$ for $\alpha > 1$ and $r > 2\alpha/(\alpha-1)$, then*

$$n^{1/2}(\mathcal{P}_n - \alpha) \Rightarrow \sigma N(0,1)$$

as $n \to \infty$.

We close this section by noting that while $\mathcal{P}_n$ is only asymptotically unbiased, it can easily be modified so that it is exactly unbiased for each $n \ge 1$. In particular, let

$$\mathcal{P}_n'' = \sum_{i=1}^{\infty}p_i\frac{\sum_{j=1}^{n}X_j(i)}{N_n(i)}\frac{I(N_n(i)\ge 1)}{1-(1-p_i)^n},$$

and note that if $E|X| < \infty$, then $E\mathcal{P}_n'' = \alpha$ for $n \ge 1$. This estimator achieves unbiasedness by heavily weighting the sample strata with extreme values of i. We leave the development of CLT's for infinitely stratified versions of this modified estimator to future work.

3. Aysmptotic Confidence Intervals for Post-Stratified Estimators

In view of Theorem 2, we can obtain a fixed sample size confidence interval procedure for α by constructing an appropriate sample variance estimator for σ^2. On $\{N_n(i) \geq 2\}$, set

$$\bar{X}(i,n) = \frac{\sum_{j=1}^n X_j I_{ij}}{N_n(i)},$$
$$s^2(i,n) = \frac{1}{N_n(i)-1} \sum_{j=1}^n (X_j - \bar{X}(i,n))^2 I_{ij}$$

and put

$$s_n^2 = \sum_{i=1}^\infty p_i s^2(i,n) I(N_n(i) \geq 2).$$

Finally, put $\eta_i \triangleq E(X_1(i) - \mu_1)^4$.

PROPOSITION 2. *Suppose that* $\sum_{i=1}^\infty p_i^2 \eta_i < \infty$ *and* $\sum_{i=1}^\infty p_i \sigma_i^2 < \infty$. *Then*

$$s_n^2 \Rightarrow \sigma^2$$

as $n \to \infty$.

Proof.

$$s_n^2 - \sigma^2 = \sum_{i=1}^\infty p_i [s^2(i,n) - \sigma_i^2] I(N_n(i) \geq 2) - \sum_{i=1}^\infty p_i \sigma_i^2 I(N_n(i) \leq 1),$$

and observe that

$$E \sum_{i=1}^\infty p_i \sigma_i^2 I(N_n(i) \leq 1) = \sum_{i=1}^\infty p_i \sigma_i^2 P(N_n(i) \leq 1) \to 0$$

as $n \to \infty$, by the Dominated Convergence Theorem. Markov's inequality then implies that $\sum_{i=1}^\infty p_i \sigma_i^2 I(N_n(i) \leq 1) \Rightarrow 0$ as $n \to \infty$.

Furthermore, for $i \neq j$,

$$\begin{aligned} &E[(s^2(i,n) - \sigma_i^2)(s^2(j,n) - \sigma_j^2) I(N_n(i) \geq 2, N_n(j) \geq 2) | \mathcal{G}_n] \\ &= I(N_n(i) \geq 2, N_n(j) \geq 2) E[(s^2(i,n) - \sigma_i^2) | \mathcal{G}_n] E[(s^2(j,n) - \sigma_j^2) | \mathcal{G}_n] \\ &= 0. \end{aligned}$$

Hence, as in the proof of Theorem 1,

$$E\left(\sum_{i=1}^\infty p_i [s^2(i,n) - \sigma_i^2] I(N_n(i) \geq 2) \right)^2 \leq \sum_{i=1}^\infty p_i^2 E(s^2(i,n) - \sigma_i^2)^2 I(N_n(i) \geq 2). \tag{11}$$

But p.348 of [Cramér(1946)] shows that there exists a constant $c < \infty$ such that

$$E[((s^2(i,n) - \sigma_i^2)^2) | \mathcal{G}_n] \leq c \eta_i / N_n(i)$$

on $\{N_n(i) \geq 2\}$. Since $\sum_{i=1}^\infty p_i^2 \eta_i < \infty$ and $E N_n(i)^{-1} I(N_n(i) \geq 2) \to 0$ as $n \to \infty$ for each $i \geq 1$, the Dominated Convergence Theorem implies that (11) converges to 0 as $n \to \infty$. Markov's inequality then shows that

$$\sum_{i=1}^\infty p_i [s^2(i,n) - \sigma_i^2] I(N_n(i) \geq 2) \Rightarrow 0$$

as $n \to \infty$, proving the result. $\square$

We are now ready to state a CLT that can be used to justify asymptotic confidence intervals for α.

THEOREM 3. *Suppose that* $0 < \sum_{i=1}^{\infty} p_i \sigma_i^2 < \infty$, $\sum_{i=1}^{\infty} p_i^2 \eta_i < \infty$, $E|\mu_J|^r < \infty$ *and* $EJ^\alpha < \infty$ *for* $\alpha > 1$ *and* $r > 2\alpha/(\alpha-1)$. *Then,*

$$n^{1/2}(\mathcal{P}_n - \alpha)/\sqrt{s_n^2} \Rightarrow N(0,1)$$

as $n \to \infty$.

The proof is immediate from Theorem 2 and Proposition 2.

Theorem 3 proves that if z is chosen so that $P(-z \le N(0,1) \le z) = 1 - \delta$, then

$$P\left(\alpha \in \left[\mathcal{P}_n - z\sqrt{\frac{s_n^2}{n}}, \mathcal{P}_n + z\sqrt{\frac{s_n^2}{n}}\right]\right) \to 1 - \delta \tag{12}$$

as $n \to \infty$. The result (12) justifies the use of

$$\left[\mathcal{P}_n - z\sqrt{\frac{s_n^2}{n}}, \mathcal{P}_n + z\sqrt{\frac{s_n^2}{n}}\right]$$

as an asymptotic $100(1-\delta)\%$ confidence interval for α.

4. Application to Debiased MLMC

We start with a brief discussion of debiased MLMC. Suppose that our goal is to compute $\alpha = EY$, where $E|Y| < \infty$ and Y is either expensive or impossible to exactly simulate. For example, Y could be some path functional associated with the solution to a stochastic differential equation (SDE). In such a setting, it is natural to seek alternative numerical algorithms that utilize approximations to Y rather than Y itself.

One means of constructing such alternatives starts with a sequence $(Y_n : n \ge 1)$ of integrable approximations to Y for which Y_n is cheaper to simulate and satisfies

$$\sum_{n=1}^{\infty} E|Y_n - Y| < \infty.$$

Set $\Delta_n = Y_n - Y_{n-1}$ for $n \ge 1$ (with $Y_0 \triangleq 0$). If N is a positive integer-valued rv independent of $(\Delta_n : n \ge 1)$, then

$$Z \triangleq \sum_{n=1}^{N} \frac{\Delta_n}{P(N \ge n)} \tag{13}$$

is unbiased for α. One can now compute α by generating iid copies of Z, and averaging them to estimate α. This is the basis of debiased MLMC; see [Rhee and Glynn(2015)].

Under suitable additional assumptions, $\mathrm{Var}(Z) < \infty$ and is given by

$$\mathrm{Var}(Z) = \sum_{n=1}^{\infty} \frac{\bar{\nu}_n}{P(N \ge n)} - \alpha^2, \tag{14}$$

where

$$\bar{\nu}_n = \|Y_{n-1} - Y\|_2^2 - \|Y_n - Y\|_2^2,$$

and $\|W\|_2 \triangleq \sqrt{E\,W^2}$ for a generic rv W. Note that the rv Z is randomized via the use of N, and this adds some unnecessary random variability to Z. It is therefore natural to consider the use of stratification as a means of eliminating some or all of this additional randomness. [Vihola(2015)] shows that when stratification is applied, the corresponding unbiased estimator has an asymptotic variance given by

$$\sum_{n=1}^{\infty} \frac{\beta_n}{P(N \geq n)}, \tag{15}$$

where

$$\beta_n = \mathrm{Var}(Y_{n-1} - Y) - \mathrm{Var}(Y_n - Y).$$

Furthermore, the variance (15) is smaller than the variance (14) associated with Z by an amount

$$\sum_{n=1}^{\infty} (EY_n - EY)^2 \left(\frac{1}{P(N \geq n+1)} - \frac{1}{P(N \geq n)} \right); \tag{16}$$

see Remark 20 of [Vihola(2015)]. The variance reduction (16) can be substantial in some cases.

The obvious strata $(A_i : i \geq 1)$ to use here correspond to setting $A_i = \{N = i\}$ for $i \geq 1$. In this case, the rv Z has the distribution of

$$Z(i) \triangleq \sum_{k=1}^{i} \frac{\Delta_k}{P(N \geq k)},$$

conditional on A_i, for $i \geq 1$. Then,

$$\sigma_i^2 = \mathrm{Var}(Z(i)) = E\left(\sum_{k=1}^{i} \frac{\tilde{\Delta}_k}{P(N \geq k)} \right)^2,$$

where $\tilde{\Delta}_k = \Delta_k - E\Delta_k$ for $k \geq 1$.

The theory of Sections 2 and 3 now applies directly to the above debiased MLMC estimator. Consequently, our theory allows one, for the first time, to construct asymptotically valid confidence intervals for stratified debiased MLMC estimators.

5. A Numerical Example

To illustrate the performance of our infinitely stratified estimators, we consider an option pricing problem in the SDE context. The underlying diffusion process obeys the SDE

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

where the parameters are $r = 0.05$, $\sigma = 0.2$ and $X(0) = 100$. We focus on computing the "final value" European call option price $e^{-rt}E(\max(X(t) - K, 0))$ with maturity $t = 1$ at various strike prices $K = 90, 100, 110$.

We implement the debiased MLMC estimator introduced in Section 4 and its stratified version, with approximating sequence $(Y_n : n \geq 1)$ constructed from the Milstein time-discretization of the SDE with step size $2^{-n}t$. The randomization N is chosen as a positive integer-valued rv with tail probabilities $P(N \geq n) = 2^{-1.25(n-1)}$ for $n \geq 1$ and we use the natural strata $A_i = \{N = i\}$ for $i \geq 1$ for the stratified debiased MLMC estimator. For each strike price, we implement both estimators for sample sizes of 1000, 4000, 16000 and 64000. Finally, in each experiment, we construct a 90% confidence interval (c.i.) for the mean based on the normal approximation, 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 C.I. 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 stratified estimator achieves a variance reduction universally in all settings, with associated variance reductions of up to 50%. (The variance reduction can be even larger when K is smaller (based on other unreported computations).) We further note that the stratified estimator demonstrates square root convergence rate, as the length of the confidence interval roughly halves each time that the sample size is multiplied by a factor of four.

Table 1. Computational Performance for Stratified and Non-Stratified Debiased MLMC

Strike	True	Sample	Stratified		Non-Stratified	
Price	Value	Size	C.I.	Coverage	C.I.	Coverage
$K = 90$	16.6994	1000	16.67 ± 0.91	89.0% ± 1.62%	16.70 ± 1.23	91.1% ± 1.48%
		4000	16.69 ± 0.45	90.1% ± 1.55%	16.69 ± 0.62	90.8% ± 1.50%
		16000	16.70 ± 0.23	88.9% ± 1.63%	16.70 ± 0.31	89.2% ± 1.61%
		64000	16.70 ± 0.11	89.6% ± 1.58%	16.70 ± 0.16	89.5% ± 1.59%
$K = 100$	10.4506	1000	10.42 ± 0.77	88.0% ± 1.69%	10.44 ± 0.94	93.7% ± 1.26%
		4000	10.44 ± 0.38	90.2% ± 1.54%	10.45 ± 0.47	91.4% ± 1.45%
		16000	10.45 ± 0.19	89.7% ± 1.58%	10.45 ± 0.23	88.9% ± 1.63%
		64000	10.45 ± 0.09	90.3% ± 1.53%	10.45 ± 0.11	90.5% ± 1.52%
$K = 110$	6.0401	1000	6.01 ± 0.61	90.1% ± 1.55%	6.03 ± 0.68	88.1% ± 1.68%
		4000	6.04 ± 0.31	88.9% ± 1.63%	6.04 ± 0.34	91.2% ± 1.47%
		16000	6.04 ± 0.15	91.0% ± 1.48%	6.04 ± 0.17	91.0% ± 1.48%
		64000	6.04 ± 0.08	89.4% ± 1.60%	6.04 ± 0.09	89.3% ± 1.60%

References

[Asmussen and Glynn(2007)] Asmussen, S., Glynn, P.W. (2007) *Stochastic Simulation.* Volume 57, Springer Science & Business Media.

[Cramér(1946)] Cramér, H. (1946) *Mathematical Methods of Statistics.* Princeton University Press.

[Bratley, Fox and Schrage(1987)] Bratley, P., Fox, B.L. and Schrage, L.E. (1987) *A Guide to Simulation.* Springer Science & Business Media.

[McLeish(2011)] McLeish, D. (2011) A general method for debiasing a Monte Carlo estimator. *Monte Carlo Meth. and Appl.* 17(4):301–315.

[Rhee and Glynn(2012)] Rhee, C.-H., Glynn, P.W. (2012) A new approach to unbiased estimation for SDE's. *Proceedings of the Winter Simulation Conference 2012*, Winter Simulation Conference.

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

[Serfling(1980)] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics.* Volume 162, John Wiley & Sons.

[Vihola(2015)] Vihola, M. (2015) Unbiased estimators and multilevel Monte Carlo. *arXiv preprint arXiv:1512.01022.*