Unbiased Estimation with Square Root Convergence for SDE Models

Chang-han Rhee
Peter W. Glynn

Stanford University
Stanford, CA, 94305, USA

Abstract

In many settings in which Monte Carlo methods are applied, there may be no known algorithm for exactly generating the random object for which an expectation is to be computed. Frequently, however, one can generate arbitrarily close approximations to the random object. We introduce a simple randomization idea for creating unbiased estimators in such a setting based on a sequence of approximations. Applying this idea to computing expectations of path functionals associated with stochastic differential equations (SDEs), we construct finite-variance unbiased estimators with a "square root convergence rate" for a general class of multi-dimensional SDEs. We then identify the optimal randomization distribution. Numerical experiments with various path functionals of continuous-time processes that often arise in finance illustrate the effectiveness of our new approach.

1 Introduction

Monte Carlo methods are powerful tools with which to study systems that are too difficult to examine analytically. In particular, typical Monte Carlo simulation methods enjoy the following pleasing properties: 1) the convergence rate is $O(c^{-1/2})$ regardless of the dimension of the problem, where c is the computational budget, and 2) the central limit theorem provides a simple mechanism for building error estimates. In many settings, however, there may be no known algorithm for exactly generating the random object for which an expectation is to be computed. In such settings, while one typically can generate arbitrarily close approximations to the random object, closer approximation generally takes more computational resources, and this often leads to a slower convergence rate. Also, the bias from the approximation error is typically much harder to estimate than the error from the variance. In this paper, we introduce a general approach to constructing unbiased estimators based on a family of such biased estimators, show how this approach applies to the setting of computing solutions of stochastic differential equations (SDEs), and illustrate the method's effectiveness with extensive numerical experiments.

We consider, in this paper, the problem of computing an expectation of the form $\alpha = \mathbf{E}f(X)$, where $X = (X(t) : t \geq 0)$ is the solution to the SDE

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t), \tag{1}$$

where $B = (B(t) : t \geq 0)$ is an m-dimensional standard Brownian motion, $\mu : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, $f : C[0,1] \to \mathbb{R}$, and $C[0,1]$ is the space of continuous functions mapping $[0,1]$ into $\mathbb{R}^d$. The functions μ and σ model respectively the state-dependent drift and volatility of X. SDEs are extensively used in mathematical finance to describe the underlying processes in financial markets; specific examples include the modeling of asset prices, interest rates, volatility, and default intensity. The expectations associated with such models are of fundamental interest for the purpose of model calibration, prediction, and pricing financial derivatives.

In general, one cannot simulate the random variable (rv) $f(X)$ exactly because it is rarely possible to generate the underlying infinite-dimensional object X exactly. However, X typically can be approximated by a discrete-time approximation $X_h(\cdot)$. For example, the simplest such approximation is the Euler discretization scheme defined by

$$X_h((j+1)h) = X_h(jh) + \mu(X_h(jh))h + \sigma(X_h(jh))(B((j+1)h) - B(jh)), \tag{2}$$

at the time points $0, h, 2h, ...$, and by (for example) linear interpolation at the intermediate time points. Note that (2) simply replaces the differentials in (1) with finite differences; the dynamics of equation (2) is only an approximation of the dynamics represented by equation (1), and hence, the random variable $f(X_h)$ generated by (2) is only an approximation of the original random variable $f(X)$. That is, $f(X_h)$ is a biased estimator of α. Although one can approximate $f(X)$ with $f(X_h)$ arbitrarily close by choosing h small enough, small h results in a large computational expense—proportional to $1/h$—for each copy of $f(X_h)$. The traditional approach to addressing this difficulty is to carefully select the step size h and the number of independent replications as a function of the computational budget c so that the errors from the bias and the variance gets balanced, so as to maximize the rate of convergence. However, such an approach inevitably leads to slower convergence rates than the canonical "square root" convergence rate $O(c^{-1/2})$ associated with typical Monte Carlo methods in the presence of unbiased finite-variance estimators; see Duffie and Glynn (1995).

In the past decade, there have been two major breakthroughs that address such difficulties. The first idea is that of exact sampling, suggested by Beskos and Roberts (2005). Their key insight is to transform a given SDE into an SDE with a unit diffusion coefficient via the Lamperti transformation so that the transformed process has a law equivalent to that of Brownian motion. Then, one can apply acceptance-rejection to sample from the exact distribution of the transformed process. One then applies the inverse of the Lamperti transformation to recover the exact sample from the original SDE. Beskos and Roberts's idea was extended to cover a more general class of SDEs by Chen and Huang (2013), and to jump diffusions, by Giesecke and Smelov (2013). Although these algorithms completely eliminate the bias from the discretization, the implementation of the algorithms requires a great deal of care and effort, and the acceptance-rejection sampling step can become inefficient when one is dealing with processes whose laws are far from that of Brownian motion. More importantly, the application is limited to scalar SDEs and to a special class of multi-dimensional SDEs because there is no analog of the Lamperti transformation for general multi-dimensional SDEs.

The second breakthrough is the multilevel Monte Carlo method introduced by Giles (2008b). By intelligently combining biased estimators with multiple step-sizes, multilevel Monte Carlo dramatically improves the rate of convergence and can even, in many settings, achieve the canonical square root convergence rate associated with unbiased Monte Carlo. This approach is not restricted to scalar processes, is much easier to implement than exact sampling algorithms, and improves the efficiency of computation by orders of magnitude for accuracies of practical relevance. However, multilevel Monte Carlo does not construct an unbiased estimator; instead, it is designed to produce an estimator with a controlled bias for the desired error tolerance.

By contrast, we show here how one can go one step further and construct unbiased estimators in a similar computational setting. The new algorithm is the first simulation algorithm that is both unbiased and achieves the square root convergence rate for multi-dimensional SDEs. We also provide a thorough discussion of the optimal choice of randomization distributions, suggest efficient algorithms for computing the optimal distributions, and numerically establish that the new randomized estimators are competitive with and often more efficient than multilevel Monte Carlo methods for typical examples that arise in finance. A preliminary result (without rigorous proof) for one of the three unbiased estimators discussed in this paper was announced in Rhee and Glynn (2012).

The remainder of this paper is organized as follows: Section 2 discusses the main randomization idea and introduces three different ways of constructing unbiased estimators. Section 3 shows how to optimize the performance of these estimators, while Section 4 discusses what can be done when we cannot achieve square root convergence. Section 5 concludes with a discussion of the implementation and describes our computational experience with the new approach.

2 A Simple Randomization Idea and Square Root Convergence

This section introduces the main idea of the paper: how one can construct unbiased estimators when only biased samplers are available, and under which conditions the new estimators can achieve the canonical square root convergence rate with respect to the computational budget.

Let L^2 be the Hilbert space of square integrable rv's, and define $\|W\|_2 \triangleq (\mathbf{E}W^2)^{1/2}$ for $W \in L^2$. Suppose now that we wish to compute an expectation of the form $\alpha = \mathbf{E}Y$, for some rv $Y \in L^2$.

We are unable to generate Y in finite (computer) time, but we assume that we have an ability to generate a sequence $(Y_n : n \geq 0)$ of L^2 approximations, each of which can be generated in finite time, for which $\|Y_n - Y\|_2 \to 0$ as $n \to \infty$ (i.e. $(Y_n : n \geq 0)$ is a sequence of L^2 rv's converging to Y in L^2). We have in mind settings in which the computational effort required to generate Y_n increases to infinity as $n \to \infty$.

Let $\Delta_n = Y_n - Y_{n-1}$ for $n \geq 0$ (with $Y_{-1} \triangleq 0$), and note that the L^2 convergence implies that $\mathbf{E}Y_n \to \mathbf{E}Y$ as $n \to \infty$. As a consequence, we can write

$$\mathbf{E}Y = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{E}\Delta_k. \tag{3}$$

The summation representation (3) for $\mathbf{E}Y$, in conjunction with a simple randomization idea, suggests a possible unbiased estimator for $\mathbf{E}Y$ that can be computed in finite time. In particular, let N be a finite-valued non-negative integer-valued rv, independent of $(Y_n : n \geq 0)$, for which $\mathbf{P}(N \geq n) > 0$ for all $n \geq 0$, and set

$$\bar{Z}_n = \sum_{k=0}^{n\wedge N} \Delta_k / \mathbf{P}(N \geq k).$$

where $a \wedge b \triangleq \min(a, b)$. Fubini's theorem applies, so that

$$\mathbf{E}\bar{Z}_n = \mathbf{E}\sum_{k=0}^{n} \frac{\Delta_k}{\mathbf{P}(N \geq k)} \mathbb{I}(N \geq k) = \sum_{k=0}^{n} \frac{\mathbf{E}\Delta_k}{\mathbf{P}(N \geq k)} \mathbf{E}\mathbb{I}(N \geq k) = \sum_{k=0}^{n} \mathbf{E}\Delta_k = \mathbf{E}Y_n,$$

so $\bar{Z}_n$ is an unbiased estimator for $\mathbf{E}Y_n$. But $\bar{Z}_n$ converges a.s. to

$$\bar{Z} = \sum_{k=0}^{N} \Delta_k / \mathbf{P}(N \geq k). \tag{4}$$

It therefore seems reasonable to expect that $\bar{Z}$ should be an unbiased estimator for $\mathbf{E}Y$, under appropriate conditions.

Theorem 1. *If*

$$\sum_{n=1}^{\infty} \frac{\|Y_{n-1} - Y\|_2^2}{\mathbf{P}(N \geq n)} < \infty, \tag{5}$$

then $\bar{Z}$ is an element of L^2, is unbiased as an estimator of $\mathbf{E}Y$, and

$$\mathbf{E}\bar{Z}^2 = \sum_{n=0}^{\infty} \bar{\nu}_n / \mathbf{P}(N \geq n), \tag{6}$$

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

The proof of Theorem 1 can be found later in this section; curiously, $\mathbf{var}\,\bar{Z}$ depends on the joint distribution of the Y_i's only through the L^2 norms of the $Y_i - Y$'s. Theorem 1 establishes that $\bar{Z}$ is an unbiased estimator for $\mathbf{E}Y$ (that clearly can be generated in finite time). This approach to constructing an unbiased estimator was previously introduced by McLeish (2011), and was later rediscovered independently by the current authors (Rhee and Glynn 2012); McLeish refers to this idea as "debiasing" the sequence $(Y_n : n \geq 0)$. The idea of introducing the random time N so as to reduce an infinite sum to a finite sum with the same expectation goes back to Glynn (1983), and perhaps even earlier. One important feature of this unbiased estimator is that unbiasedness is achieved in a setting where very little needs to be known a priori regarding the exact form of the bias. In particular, we are not assuming here a parametric functional form for the bias of Y_n as an estimator for α, and then attempting to estimate the unknown parameters from the sample. So, this methodology is potentially broadly applicable (to settings well beyond the SDE context that is the focus of this paper).

Under the conditions of Theorem 1, asymptotically valid confidence intervals for $\mathbf{E}Y$ can easily be computed, along well-known lines. Specifically, for a given sample size m that is large, suppose that one generates m iid replicates $\bar{Z}(1),\ldots,\bar{Z}(m)$ of the rv $\bar{Z}$, and computes

$$\bar{\alpha}_m \triangleq 1/m \sum_{i=1}^{m} \bar{Z}(i),$$

$$s_m \triangleq \sqrt{1/(m-1) \sum_{i=1}^{m} (\bar{Z}(i) - \bar{\alpha}_m)^2}.$$

An approximate $100(1-\delta)$ percent confidence interval for $\mathbf{E}Y$ is given by $[L_m, R_m]$, where $L_m = \bar{\alpha}_m - z s_m m^{-1/2}$ and $R_m = \bar{\alpha}_m + z s_m m^{-1/2}$. As usual, z is chosen so that $\mathbf{P}(-z \leq N(0,1) \leq z) = 1 - \delta$, where $N(0,1)$ denotes a normal rv with mean zero and unit variance. Provided that $0 < \mathbf{var}\,\bar{Z} < \infty$, $\mathbf{P}(\mathbf{E}Y \in [L_m, R_m]) \to 1 - \delta$ as $m \to \infty$. This algorithm can also be implemented in a sequential mode, in which one first sets an appropriate desired error tolerance ϵ, and samples iid $\bar{Z}$ replicates until the first time $K(\epsilon)$ at which $(L_{K(\epsilon)} - R_{K(\epsilon)}) < 2\epsilon$. Under our assumption that $0 < \mathbf{var}\,\bar{Z} < \infty$, a slightly modified version of this algorithm can be shown to be asymptotically valid, in the sense that $\mathbf{P}(\mathbf{E}Y \in [L_{K(\epsilon)}, R_{K(\epsilon)}]) \to 1 - \delta$, as $\epsilon \to 0$; see Glynn and Whitt (1992b) for details.

In contrast to McLeish (2011), our interest in this paper is in explaining the profound consequences of this randomization idea in the setting of SDEs. In the SDE context, the most natural means of constructing an approximation Y_0 to Y is by running a time-discretization algorithm with a single time step, so that $h = 1$ in the notation of Section 1. The n'th approximation is then obtained by doubling the number of time steps relative to the $(n-1)$'st such approximation, so that Y_n is the time-discretization associated with time-step increment $h = 2^{-n}$. In the conventional application of such discretization schemes, we fix a value of n and generate independent and identically distributed (iid) copies of Y_n as a means of computing a (biased) estimator of $\mathbf{E}Y$. In such an implementation, the joint distribution of the Y_n's (and indeed even whether Y_n is jointly distributed with Y) is immaterial to the algorithmic implementation; only the marginal distribution of Y_n affects the algorithm. In contrast, our estimator $\bar{Z}$ can be constructed only when the sequence $(Y_n : 0 \leq n \leq N)$ can be jointly generated. (We will discuss a relaxation of this requirement in Section 5). Furthermore, it is critical that we construct a simulatable joint distribution for the Y_n's for which (5) is valid. In the language of probability, a key algorithmic element here is the choice of "coupling" (i.e. joint probability law) between the Y_n's that is utilized.

Fortunately, it is easy in the SDE context, given a specific discretization scheme, to build a good coupling. In particular, we can conceptually take the view that there is a single Brownian motion B that drives both the SDE (1) and all its discretizations. From this perspective, given N, one first generates the Brownian time increments associated with the finest time discretization, namely $h = 2^{-N}$, and computes Y_N. To generate the approximation Y_{N-1}, one sums the $2j$'th and $(2j-1)$'st increments together, thereby obtaining the j'th increment needed by the $(N-1)$'st approximation, namely $B(j2^{-(N-1)}) - B((j-1)2^{-(N-1)})$. By summing successive pairs of increments together and applying the discretization scheme, one now obtains Y_{N-1}. A similar procedure of summing pairs of Brownian increments from the approximation Y_i, followed by applying the chosen discretization scheme to the newly combined increments, leads to approximation Y_{i-1}, so that $Y_N, Y_{N-1}, Y_{N-2}, \ldots$, Y_2, Y_1, Y_0 can be generated (in that order). (If one wishes to build the approximations in the order Y_0, Y_1,..., one would use a "Brownian bridge" simulation scheme to successively refine the discretization; see Section 2 of Rhee and Glynn (2012) for details). As we shall see later, one can then frequently argue that these approximations are such that

$$\|Y_n - Y\|_2 = O(2^{-np}), \tag{7}$$

for some $p > 0$, as $n \to \infty$, where $O(a_n)$ is a sequence that is bounded by a constant multiple of $|a_n|$. The parameter p reflects what is known in the SDE numerical computation literature as the strong order of the scheme. Of course, in the presence of (7), it is easy to construct distributions for N that satisfy (5). Further discussion of condition (7) in the setting of SDEs can be found in Section 5.

In a conventional implementation of a p'th (weak) order discretization scheme, one chooses the time step h (that controls the bias of the estimator) and the number of iid replicates n (controlling the variance of the estimator), and optimally chooses n and h so as to minimize the resulting mean square error of the estimator. For a given computer time budget c, the fastest rate of convergence that can be achieved is of order $c^{-\frac{p}{2p+1}}$ (which arises when h is chosen to be of order $h = c^{-\frac{1}{2p+1}}$); see Duffie and Glynn (1995) for details. Thus, conventional implementations of SDE schemes always lead to rates of convergence that are "subcanonical", in the sense that the rate of convergence is slower than the $c^{-1/2}$ rate that is often exhibited in the Monte Carlo setting. Of course, the higher order a scheme one implements (so that p is larger), the closer one can get to the canonical $c^{-1/2}$ rate.

We shall now argue that use of the estimator $\bar{Z}$ can dramatically change the situation. In particular, for any $p > 1/2$, our randomization idea easily leads to unbiased estimators that can achieve the canonical convergence rate of $c^{-1/2}$. Thus, not only is the convergence rate improved (to the canonical "square root" rate) by applying this simple idea, but there is no compelling reason for implementing (very) high order schemes, because the canonical rate can already be achieved once $p > 1/2$. This is an important observation, as high-order schemes are complicated to implement, and typically involve a very high computational cost per time step (because many partial derivatives of the drift and volatility functions need to be computed at each time step); see Kloeden and Platen (1992) for details. It should be noted that McLeish (2011) does not explore this connection between debiasing and its ability to modify the rate of convergence for a p'th order scheme from $c^{-\frac{p}{2p+1}}$ to $c^{-1/2}$, and does not construct efficient estimators (with square root convergence) for SDEs, whereas this is the focus of the current paper.

To study the rate of convergence for $\bar{\alpha}_n$, we need to take into account the computer time $\bar{\tau}$ required to generate each $\bar{Z}$. If $\bar{t}_j$ is the expected incremental effort required to calculate Y_j,

$$\mathbf{E}\bar{\tau} = \mathbf{E}\sum_{j=0}^{N} \bar{t}_j = \sum_{j=0}^{\infty} \bar{t}_j \mathbf{P}(N \geq j). \tag{8}$$

A natural computational model in the SDE setting is to presume that the computational effort required to calculate Y_j is of order 2^j, so that in the SDE context, we set $t_j = 2^j$. For a given computational budget c, we let $\Gamma(c)$ equal the number of replicates $\bar{Z}(i)$'s of $\bar{Z}$ generated in c units of computer time, so that $\Gamma(c) = \max\{n \geq 0 : \sum_{i=1}^{n} \bar{\tau}(i) \leq c\}$ where $\bar{\tau}(i)$ denotes the required computer time to generate each $\bar{Z}(i)$. In this computational context, it is clear that the $(\bar{Z}(i), \bar{\tau}(i))$'s are iid pairs, while within each pair, $\bar{\tau}(i)$ is generally highly correlated with $\bar{Z}(i)$. Hence, $\Gamma(c)$ is a renewal counting process, and the estimator available after c units of computer time have been expended is $\bar{\alpha}(c) \triangleq \bar{\alpha}_{\Gamma(c)}$ (with the estimator defined to be equal to 0 if $\Gamma(c) = 0$). Glynn and Whitt (1992a) prove that if $\mathbf{E}\bar{\tau} < \infty$ and $\mathbf{var}\,\bar{Z} < \infty$, then

$$c^{1/2}(\bar{\alpha}(c) - \mathbf{E}Y) \Rightarrow (\mathbf{E}\bar{\tau} \cdot \mathbf{var}\,\bar{Z})^{1/2} N(0,1) \tag{9}$$

as $c \to \infty$. Thus, if we can find a distribution for N for which both (5) and (8) are finite, (9) guarantees that our randomized estimator achieves the canonical square root convergence rate. In the presence of (7), $\bar{\nu}_n = O(2^{-2np})$ and $\bar{t}_n = 2^n$; when $p > 1/2$, a choice for the distribution of N that achieves the required finiteness of both (5) and (8) is to choose N, for example, so that $\mathbf{P}(N \geq n) = 2^{-rn}$, where $1 < r < 2p$. This verifies our earlier claim that the use of the randomized estimator $\bar{Z}$ can transform a p'th order SDE scheme from one that exhibits a subcanonical rate to one that can achieve a canonical "square root" rate. Of course, we can further tune the distribution of N so that the product $\mathbf{E}\bar{\tau} \cdot \mathbf{var}\, \bar{Z}$ is minimized; we will return to this topic in Section 3. In addition, there is a question of what can be achieved in terms of the convergence rate when the order $p \in (0, 1/2]$; this will be addressed in Section 4.

The assumption that $\mathbf{E}\bar{\tau} < \infty$ is equally as important as obtaining a finite variance unbiased estimator in building a computational method that achieves the canonical rate. However, it should be emphasized that the theoretical validity of the confidence interval and sequential methodology described above does not require the finiteness of $\mathbf{E}\bar{\tau}$; only the finiteness of $\mathbf{var}\, \bar{Z}$ is needed (for which much more flexibility in choosing the distribution N is available). We further note that the improved convergence rate obtained here builds on the fact that the discretization scheme is simultaneously implemented at various discretization levels $h = 2^{-k}$, $1 \leq k \leq N$, all simulated using a common Brownian motion B. In view of the multiple levels of discretization used, it will come as no surprise that our unbiased randomized estimator is closely related to multi-level Monte Carlo (MLMC) methods; this connection is discussed further in Sections 4 and 5.

To complete this section, we introduce two new additional randomized estimators that offer similar advantages to what can be achieved by $\bar{Z}$; these estimators were not discussed in McLeish (2011). The second estimator requires choosing N so that $p_n \triangleq \mathbf{P}(N = n) > 0$ for $n \geq 0$, and setting

$$Z = \Delta_N / p_N; \tag{10}$$

in view of (7), it is easily verified that Z is unbiased as an estimator for $\mathbf{E}Y$. Furthermore, the variance can easily be computed from its second moment

$$\mathbf{E}Z^2 = \sum_{n=0}^{\infty} \mathbf{E}\Delta_n^2 / p_n;$$

the time τ required to generate Z is given by t_N, where t_n is the time required to generate Δ_n. In the SDE context, $\|\Delta_n\|_2$ is of the order of 2^{-np} when (7) is in force, and t_n is of the order of 2^n. If $\alpha(c)$ is the estimator available after expending c units of computer time to generate iid copies of Z, Glynn and Whitt (1992a) again applies if $\mathbf{E}\tau < \infty$ and $\mathbf{var}\, Z < \infty$, yielding the central limit theorem (CLT)

$$c^{1/2}(\alpha(c) - \mathbf{E}Y) \Rightarrow (\mathbf{E}\tau \cdot \mathbf{var}\, Z)^{1/2} N(0, 1) \tag{11}$$

as $c \to \infty$; we call Z the *single-term estimator* to differentiate this estimator from $\bar{Z}$, which we henceforth refer to as the *coupled-sum estimator*.

Our final estimator takes advantage of the fact that (3) continues to hold for any sequence $(\tilde{\Delta}_n : n \geq 0)$ for which $\tilde{\Delta}_n = \tilde{Y}_n - \tilde{Y}'_{n-1}$, where $(\tilde{Y}_n, \tilde{Y}'_{n-1})$ has the same marginal distribution as (Y_n, Y_{n-1}) for each $n \geq 0$. One such sequence $(\tilde{\Delta}_n : n \geq 0)$ is that in which the $\tilde{\Delta}_n$'s are independent. When we generate the $\tilde{\Delta}_n$'s in this way, we can now apply the same randomization trick used for constructing $\bar{Z}$, thereby yielding a new estimator

$$\tilde{Z} = \sum_{n=0}^{N} \tilde{\Delta}_n / \mathbf{P}(N \geq n); \tag{12}$$

we call $\tilde{Z}$ the *independent-sum estimator.*

Theorem 2. *If (5) holds, then $\tilde{Z}$ is an element of L^2 and is an unbiased estimator for $\mathbf{E}Y$. Furthermore,*

$$E\tilde{Z}^2 = \sum_{n=0}^{\infty} \tilde{\nu}_n / \mathbf{P}(N \geq n), \tag{13}$$

where $\tilde{\nu}_n = \mathbf{var}\,(Y_n - Y_{n-1}) + (\mathbf{E}Y - \mathbf{E}Y_{n-1})^2 - (\mathbf{E}Y - \mathbf{E}Y_n)^2$.

The proof can be found below. As for the coupled-sum and single-term estimators, we can again appeal to Glynn and Whitt (1992a) to understand the behavior of the independent sum estimator as a function of the computational budget c. In particular, if $\tilde{\alpha}(c)$ is the estimator available after expending c units of computer time to generate iid copies of $\tilde{Z}$,

$$c^{1/2}(\tilde{\alpha}(c) - \mathbf{E}Y) \Rightarrow (\mathbf{E}\tilde{\tau} \cdot \mathbf{var}\,\tilde{Z})^{1/2} N(0,1) \tag{14}$$

as $c \to \infty$, provided that $\mathbf{E}\tilde{\tau} < \infty$ and $\mathbf{var}\,\tilde{Z} < \infty$, where $\tilde{\tau}$ is the time required to generate $\tilde{Z}$. As for the coupled sum estimator, $\tilde{\tau}$ is of order 2^N, and $\tilde{\nu}_n$ is of order 2^{-2np} in the SDE setting, provided that (7) is in force.

Proof of Theorem 1. Put $\delta_k = Y_k - Y$, let $\rho_0 = 0$, and set $\rho_k = \inf\{j > \rho_{k-1} : \|\delta_j\|_2 \leq \|\delta_{\rho_{k-1}}\|_2\}$ for $k \geq 1$. By construction, $\rho_k \to \infty$ and $\|\delta_{\rho_k}\|_2 \leq \|\delta_j\|_2$ for $j \leq \rho_k$. We start by showing that $(\bar{Z}_{\rho_k} : k \geq 0)$ is a Cauchy sequence in L^2 whenever (5) is valid. Put $\bar{Z}'_k = \bar{Z}_{\rho_k}$ and note that if $n > m$ then

$$\bar{Z}'_n - \bar{Z}'_m = \sum_{i=\rho_m+1}^{\rho_n} \Delta_i \mathbb{I}(N \geq i) / \mathbf{P}(N \geq i)$$

and

$$(\bar{Z}'_n - \bar{Z}'_m)^2 = \sum_{i=\rho_m+1}^{\rho_n} \Delta_i^2 \mathbb{I}(N \geq i) / \mathbf{P}(N \geq i)^2 + 2 \sum_{i=\rho_m+1}^{\rho_n} \sum_{j=i+1}^{\rho_n} \frac{\Delta_i \Delta_j \mathbb{I}(N \geq j)}{\mathbf{P}(N \geq i)\mathbf{P}(N \geq j)}.$$

The independence of N from the Δ_i's implies that

$$
\begin{aligned}
\|\bar{Z}'_n - \bar{Z}'_m\|_2^2 &= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\Delta_i^2/\mathbf{P}(N \geq i) + 2 \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\Delta_i \sum_{j=i+1}^{\rho_n} \Delta_j/\mathbf{P}(N \geq i) \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[\Delta_i^2 + 2\Delta_i(Y_{\rho_n} - Y_i)]/\mathbf{P}(N \geq i).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\Delta_i^2 + 2\Delta_i(Y_{\rho_n} - Y_i) &= ((Y_i - Y_{\rho_n}) - (Y_{i-1} - Y_{\rho_n}))^2 - 2((Y_i - Y_{\rho_n}) - (Y_{i-1} - Y_{\rho_n}))(Y_i - Y_{\rho_n}) \\
&= (Y_{i-1} - Y_{\rho_n})^2 - (Y_i - Y_{\rho_n})^2 \qquad (15) \\
&\leq (Y_{i-1} - Y_{\rho_n})^2 = (\delta_{i-1} - \delta_{\rho_n})^2 \leq 2\delta_{i-1}^2 + 2\delta_{\rho_n}^2.
\end{aligned}
$$

Because of the way in which ρ_n was chosen,

$$
\|\bar{Z}'_n - \bar{Z}'_m\|_2^2 \leq 2 \sum_{i=\rho_m+1}^{\rho_n} \frac{\|\delta_{i-1}\|_2^2}{\mathbf{P}(N \geq i)} + 2 \sum_{i=\rho_m+1}^{\rho_n} \frac{\|\delta_{\rho_n}\|_2^2}{\mathbf{P}(N \geq i)} \leq 4 \sum_{i=\rho_m+1}^{\rho_n} \frac{\|\delta_{i-1}\|_2^2}{\mathbf{P}(N \geq i)} \leq 4 \sum_{i=\rho_m+1}^{\infty} \frac{\|\delta_{i-1}\|_2^2}{\mathbf{P}(N \geq i)}.
$$

In view of (5), we can make the last sum as small as we wish by choosing m sufficiently large, thereby proving that $(\bar{Z}'_n : n \geq 0)$ is Cauchy. Hence, there exists $\bar{Z}' \in L^2$ for which $\bar{Z}'_n \to \bar{Z}'$ in L^2. But recall that $\bar{Z}_n \to \bar{Z}$ a.s. as $n \to \infty$. This implies that $\bar{Z}'_n \to \bar{Z}$ in L^2. As a consequence, $\mathbf{E}\bar{Z}'_n = \mathbf{E}Y_{\rho_n} \to \mathbf{E}\bar{Z}$, proving that $\mathbf{E}\bar{Z} = \mathbf{E}Y$ and establishing the unbiasedness of $\bar{Z}$ as a estimator of $\mathbf{E}Y$.

Furthermore, the L^2 convergence of $\bar{Z}'_n$ to $\bar{Z}$ implies that $\mathbf{E}\bar{Z}'^2_n \to \mathbf{E}\bar{Z}^2$ as $n \to \infty$. The same calculation as that leading to (15) shows that

$$
\mathbf{E}\bar{Z}'^2_n = \sum_{i=0}^{\rho_n} \mathbf{E}[(Y_{i-1} - Y_{\rho_n})^2 - (Y_i - Y_{\rho_n})^2]/\mathbf{P}(N \geq i).
$$

But $\|Y_{i-1} - Y_{\rho_n}\|_2^2 = \|\delta_{i-1} - \delta_{\rho_n}\|_2^2 \leq \|\delta_{i-1}\|_2^2 + 2\|\delta_{i-1}\|_2\|\delta_{\rho_n}\|_2 + \|\delta_{\rho_n}\|_2^2 \leq 4\|\delta_{i-1}\|_2^2$, because of our choice of the subsequence $(\rho_n : n \geq 0)$. Because $\|Y_{i-1} - Y_{\rho_n}\|_2^2 \to \|Y_{i-1} - Y\|_2^2$ as $n \to \infty$, the dominated convergence theorem implies that

$$
\mathbf{E}\bar{Z}'^2_n \to \sum_{i=0}^{\infty} \mathbf{E}[(Y_{i-1} - Y)^2 - (Y_i - Y)^2]/\mathbf{P}(N \geq i),
$$

thereby verifying our expression for $\mathbf{E}\bar{Z}^2$. □

Note that Theorem 1 provides a straightforward sufficient condition for the validity of the estimator $\bar{Z}$, and a clean expression for its variance (Theorem 2.1 of McLeish (2011) requires verifying that three different sequences converge in L^2). It should also be noted that Theorem 1's key hypothesis (5) involves only squared L^2 norms, rather than the (much larger) L^2 norms themselves.

Proof of Theorem 2. The proof is similar to that of Theorem 1. Using exactly the same subsequence as that specified in the proof of Theorem 1, note that

$$
\begin{aligned}
&\left\| \sum_{i=0}^{\rho_n} \tilde{\Delta}_i \frac{\mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} - \sum_{i=0}^{\rho_m} \tilde{\Delta}_i \frac{\mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} \right\|_2^2 \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\tilde{\Delta}_i^2/\mathbf{P}(N \geq i) + 2 \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\tilde{\Delta}_i \sum_{j=i+1}^{\rho_n} \tilde{\Delta}_j/\mathbf{P}(N \geq i) \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\Delta_i^2/\mathbf{P}(N \geq i) + 2 \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}\Delta_i \mathbf{E} \sum_{j=i+1}^{\rho_n} \Delta_j/\mathbf{P}(N \geq i) \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[\Delta_i^2 + 2\Delta_i \mathbf{E}(Y_{\rho_n} - Y_i)]/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} (\|\delta_i - \delta_{i-1}\|_2^2 + 2\|\delta_i - \delta_{i-1}\|_2 \|\delta_{\rho_n} - \delta_i\|_2)/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} (2\|\delta_i\|_2^2 + 2\|\delta_{i-1}\|_2^2 + 2(\|\delta_i\|_2 + \|\delta_{i-1}\|_2)(\|\delta_{\rho_n}\|_2 + \|\delta_i\|_2))/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} (2\|\delta_i\|_2^2 + 2\|\delta_{i-1}\|_2^2 + 4(\|\delta_i\|_2 + \|\delta_{i-1}\|_2)\|\delta_i\|_2)/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} (2\|\delta_i\|_2^2 + 2\|\delta_{i-1}\|_2^2 + 2\|\delta_{i-1}\|_2^2 + 6\|\delta_i\|_2^2)/\mathbf{P}(N \geq i) \\
&= \sum_{i=\rho_m+1}^{\rho_n} (4\|\delta_{i-1}\|_2^2 + 8\|\delta_i\|_2^2)/\mathbf{P}(N \geq i).
\end{aligned}
$$

Because $\mathbf{P}(N \geq i)$ is decreasing, $\sum_{i=0}^{\infty} \|\delta_i\|_2^2/\mathbf{P}(N \geq i)$ is guaranteed to be finite when (5) is in force. The above bound therefore verifies the Cauchy property along the subsequence $(\rho_n : n \geq 0)$, establishing the unbiasedness of $\tilde{Z}$. The formula for $\mathbf{E}\tilde{Z}^2$ follows from straightforward algebraic manipulations similar to those used in the proof of Theorem 1. □

3 The Optimal Distribution for N

Our goal in this section is to discuss good choices for the distribution of N, in settings where the "work-variance products" associated with the limiting normal distributions in the CLT's (9), (11), (14) are all finite (and for which square-root convergence ensues). We start with the coupled-sum estimator. Set $\bar{\beta}_0 = \bar{\nu}_0 - \alpha^2$, $\bar{\beta}_n = \bar{\nu}_n$ for $n \geq 1$, and $\bar{F}_n = \mathbf{P}(N \geq n)$. To maximize the efficiency of the estimator $\bar{\alpha}(c)$, we need to find a distribution for N that solves the following optimization problem:

$$\begin{aligned}
\min_{\bar{F}} \quad & \bar{g}(\bar{F}) \triangleq \left(\sum_{n=0}^{\infty} \bar{\beta}_n / \bar{F}_n\right) \left(\sum_{n=0}^{\infty} \bar{t}_n \bar{F}_n\right) \\
\text{s/t} \quad & \bar{F}_i \geq \bar{F}_{i+1}, \quad \forall i \geq 0 \\
& \bar{F}_i > 0, \quad \forall i \geq 0 \\
& \bar{F}_0 = 1.
\end{aligned} \tag{16}$$

Proposition 1. *Suppose that $(\bar{\beta}_i : i \geq 0)$ is a non-negative sequence. Then,*

$$\bar{g}(\bar{F}) \geq \left(\sum_{j=0}^{\infty} \sqrt{\bar{\beta}_j \bar{t}_j}\right)^2 = \bar{g}(y^*)$$

for any $\bar{F}$ that is feasible for (16), where $y^* = (y_i^* : i \geq 0)$ is given by

$$y_i^* = \frac{\sqrt{\bar{\beta}_i / \bar{t}_i}}{\sqrt{\bar{\beta}_0 / \bar{t}_0}}.$$

If y^* is feasible for (16), y^* is a minimizer of (16).

Proof. If $\sum_{i=1}^{\infty} \bar{t}_i \bar{F}_i = \infty$, $\bar{g}(\bar{F}) = \infty$, so obviously $\bar{g}(\bar{F}) \geq \bar{g}(y^*)$. If $\sum_{i=0}^{\infty} \bar{t}_i \bar{F}_i < \infty$, set $p_i = \bar{t}_i \bar{F}_i / \sum_{j=0}^{\infty} \bar{t}_j \bar{F}_j$ for $i \geq 0$, and observe that $(p_i : i \geq 0)$ is a probability mass function. The Cauchy-Schwarz inequality implies that

$$\sum_{j=0}^{\infty} p_j \left(\sqrt{\frac{\bar{\beta}_j}{\bar{t}_j}} \frac{1}{\bar{F}_j}\right)^2 \geq \left(\sum_{i=0}^{\infty} p_j \sqrt{\frac{\bar{\beta}_j}{\bar{t}_j}} \frac{1}{\bar{F}_j}\right)^2.$$

However, this is easily seen to be equivalent to the inequality $\bar{g}(\bar{F}) \geq \bar{g}(y^*)$, proving the result. □

Note that the non-negativity of $(\bar{\beta}_n : n \geq 0)$ is equivalent to requiring that $(\|Y_n - Y\|_2^2 : n \geq 0)$ is a non-increasing sequence, and $\mathbf{var}\, Y \geq \|Y - Y_0\|_2^2$; this seems a reasonable condition that will naturally arise in some problem settings. If the sequence $(\|Y_n - Y\|_2^2 : n \geq 0)$ is not already decreasing, we can always (easily) select a subsequence $(n_k : k \geq 0)$ for which $(\|Y_{n_k} - Y\|_2^2 : k \geq 0)$ is decreasing, and use this subsequence in place of the original sequence of approximations. (In the presence of empirical data, one can estimate the magnitude of these squared norms from sample data, using the approximations at the finest level of discretization in place of Y). Of course, there is a question as to whether one could potentially lose efficiency by passing to such a subsequence, in the sense of a possible adverse impact on the work-variance product $\mathbf{E}\bar{\tau} \cdot \mathbf{var}\, \bar{Z}$.

However, it turns out that one can always decrease the work-variance product by passing to such a monotone subsequence. In particular, if $\bar{\beta}_i < 0$ for a feasible $\bar{F}$, with $\bar{F}_i > 0$, we can reduce $\bar{g}(\bar{F})$ by setting $\bar{F}_i$ to $\bar{F}_{i+1}$, and leaving $(\bar{F}_j : j \neq i)$ unchanged. This moves any "mass" in $\bar{F}$ at i to $i-1$, and effectively "collapses" the two differences $Y_i - Y_{i-1}$ and $Y_{i+1} - Y_i$ to a single difference $Y_{i+1} - Y_{i-1}$ with the newly randomized distribution. In general, $\bar{g}(\bar{F})$ is reduced by collapsing all the differences associated with i's for which $\bar{\beta}_i < 0$, and moving the mass to smaller j's for which $\bar{\beta}_j > 0$, thereby modifying the objective function to a new $\bar{g}(\cdot)$ in which all the $\bar{\beta}_k$'s are positive. Thus, there is always a "canonical" monotone sequence to which one can pass that guarantees a reduction in the work-variance product. So, non-negativity of $(\bar{\beta}_k : k \geq 0)$ (or, equivalently, the monotonicity of the squared norms) can essentially be assumed, without loss of generality.

Returning to Proposition 1, we note that it provides an optimal distribution for N when y^* is feasible for (16). In the applications that we have in mind, the sequence $(\bar{t}_n : n \geq 0)$ will be non-decreasing. It follows that if $(\bar{\beta}_n : n \geq 0)$ is decreasing, then y^* is feasible. But the assumption that the $\bar{\beta}_n$'s are decreasing is precisely equivalent to requiring that the sequence $(\|Y_n - Y\|_2^2 : n \geq 0)$ be convex (i.e. $\frac{1}{2}(\|Y_{n-1} - Y\|_2^2 + \|Y_{n+1} - Y\|_2^2) \geq \|Y_n - Y\|_2^2$ for $n \geq 1$) and $\frac{1}{2}(\mathbf{var}\, Y + \|Y_1 - Y\|_2^2) \geq \|Y_0 - Y\|_2^2$. As in our discussion of monotonicity, we can always choose to pass to a decreasing convex subsequence of the $\|Y_n - Y\|_2^2$'s. However, in this setting there is no canonical "convexification" to which one can pass that always guarantees a reduction in the work-variance product. So, convexifying our sequence $(Y_n : n \geq 0)$ may result in a loss of efficiency. As a consequence, we will now consider the solution of (16) when y^* is infeasible.

Proposition 2. *Suppose that $(\bar{\beta}_i : i \geq 0)$ is a positive sequence and that the $\bar{t}_n$'s are bounded below by a positive constant. Then, (16) achieves its minimum over the feasible region.*

Proof. If $\bar{g}(\bar{F}) = \infty$ for all feasible $\bar{F}$, the result is trivial. If $\bar{g}(\bar{F}) < \infty$ for some feasible $\bar{F}$, let $(\bar{F}^{(k)} : k \geq 0)$ be a sequence of feasible solutions for which $\bar{g}(F^{(k)}) \to g_*$, where $g_* < \infty$ is the infimum of $\bar{g}(\cdot)$ over the feasible region. For $k \geq k_0$,

$$\bar{g}(F^{(k)}) \leq g_* + 1,$$

so

$$\bar{\beta}_0 \cdot \sum_{j=0}^{\infty} \bar{t}_j \bar{F}_j^{(k)} \leq g_* + 1.$$

It follows from Markov's inequality that

$$\bar{F}_n^{(k)} \leq \frac{\sum_{j=0}^{\infty} \bar{t}_j \bar{F}_j^{(k)}}{\sum_{j=0}^{n} \bar{t}_j} \leq \frac{g_* + 1}{\bar{\beta}_0 \cdot \sum_{j=0}^{n} \bar{t}_j}$$

for $k \geq k_0$. Since the $\bar{t}_j$'s are bounded below by a positive constant, $\sum_{j=0}^{n} \bar{t}_j \to \infty$ as $n \to \infty$, so evidently the distributions corresponding to $(\bar{F}^{(k)} : k \geq k_0)$ are tight. As a consequence, Prohorov's theorem (see, for example, Billingsley 1999) guarantees that there exists a subsequence $(k_n : n \geq 0)$ and $\bar{F}^*$ for which

$$\bar{F}_j^{(k_n)} \to \bar{F}_j^*$$

as $n \to \infty$, for each $j \geq 0$. Fatou's lemma then yields

$$\bar{g}(\bar{F}^*) \leq \liminf_{n\to\infty} \bar{g}(\bar{F}^{(k_n)}) = g_*. \tag{17}$$

Since $\bar{g}(\bar{F}^*) < \infty$, $\bar{F}_i^* > 0$ for $i \geq 0$, and hence $\bar{F}^*$ is feasible for (16). The inequality (17) therefore implies that $\bar{F}^*$ is an optimal solution of (16). □

Let $i_0^* = 0$ and $i_j^* = \inf\{k > i_{j-1}^* : \bar{F}_k < \bar{F}_{i_{j-1}^*}\}$ for $j \geq 1$, so that the $(i_j^* - 1)$'s are the integers upon which the optimal distribution for N is supported. For a generic strictly increasing integer-valued sequence $J = (i_j : j \geq 0)$ for which $i_0 = 0$, let

$$\bar{\beta}_j(J) = \sum_{k=i_j}^{i_{j+1}-1} \bar{\beta}_k \qquad \text{and} \qquad \bar{t}_j(J) = \sum_{k=i_j}^{i_{j+1}-1} \bar{t}_k.$$

If $J^* = (i_j^* : j \geq 0)$, it is evident that

$$\bar{g}(\bar{F}^*) = \sum_{k=0}^{\infty} \bar{\beta}_k(J^*)/\bar{F}_{i_k^*}^* \cdot \sum_{k=0}^{\infty} \bar{t}_k(J)\bar{F}_{i_k^*}^*. \tag{18}$$

Proposition 3. *If $(\bar{\beta}_n : n \geq 0)$ and $(\bar{t}_n : n \geq 0)$ are positive sequences, then $(\bar{\beta}_k(J^*)/\bar{t}_k(J^*) : k \geq 0)$ is a strictly decreasing sequence.*

Proof. We show that if there exists k for which $\bar{\beta}_k(J^*)/\bar{t}_k(J^*) \geq \bar{\beta}_{k-1}(J^*)/\bar{t}_{k-1}(J^*)$, then $\bar{g}(\bar{F}^*)$ can be strictly decreased while maintaining feasibility, contradicting the optimality of $\bar{F}^*$. For $x \in \mathbb{R}$ and $a \geq 0$, let

$$\bar{F}_j^x = \begin{cases} \bar{F}_j^* & , j \notin \{i_{k-1}^*, \ldots, i_{k+1}^* - 1\} \\ \bar{F}_j^* - xa & , j \in \{i_{k-1}^*, \ldots, i_k^* - 1\} \\ \bar{F}_j^* + x & , j \in \{i_k^*, \ldots, i_{k+1}^* - 1\} \end{cases}.$$

Observe that $\bar{F}^x = (\bar{F}_j^x : j \geq)$ is feasible for $|x|$ sufficiently small. Set $f(x) = \bar{g}(\bar{F}^x)$ and note that

$$f(x) = \left(v + \frac{\bar{\beta}_{k-1}(J^*)}{f_{k-1} - xa} + \frac{\bar{\beta}_k(J^*)}{f_k + x}\right)\Big(w + \bar{t}_{k-1}(J^*)(f_{k-1} - xa) + \bar{t}_k(J^*)(f_k + x)\Big),$$

where

$$
\begin{aligned}
v &= \sum_{j \neq k-1,k} \bar{\beta}_j(J^*)/\bar{F}^*_{i^*_j},\\
w &= \sum_{j \neq k-1,k} \bar{t}_j(J^*)\bar{F}^*_{i^*_j},\\
f_j &= \bar{F}^*_{i^*_j}, j \geq 0.
\end{aligned}
$$

Then,

$$
\begin{aligned}
f'(0) = &\left(a\frac{\bar{\beta}_{k-1}(J^*)}{f^2_{k-1}} - \frac{\bar{\beta}_k(J^*)}{f^2_k}\right)w + \Big(\bar{t}_k(J^*) - \bar{t}_{k-1}(J^*)a\Big)v\\
&\quad + a\bar{t}_{k-1}(J^*)\bar{t}_k(J^*)\left(\frac{\bar{\beta}_{k-1}(J^*)}{\bar{t}_{k-1}(J^*)}\frac{f_k}{f^2_{k-1}} - \frac{\bar{\beta}_k(J^*)}{\bar{t}_k(J^*)}\frac{1}{f_k}\right)\\
&\quad + \bar{t}_{k-1}(J^*)\bar{t}_k(J^*)\left(\frac{\bar{\beta}_{k-1}(J^*)}{\bar{t}_{k-1}(J^*)}\frac{1}{f_{k-1}} - \frac{\bar{\beta}_k(J^*)}{\bar{t}_k(J^*)}\frac{f_{k-1}}{f^2_k}\right)\\
&< \Big(a\bar{\beta}_{k-1}(J^*) - \bar{\beta}_k(J^*)\Big)\frac{w}{f^2_k} + \Big(\bar{t}_k(J^*) - \bar{t}_{k-1}(J^*)a\Big)v\\
&\quad + a\frac{\bar{t}_{k-1}(J^*)\bar{t}_k(J^*)}{f_k}\left(\frac{\bar{\beta}_{k-1}(J^*)}{\bar{t}_{k-1}(J^*)} - \frac{\bar{\beta}_k(J^*)}{\bar{t}_k(J^*)}\right)\\
&\quad + \frac{\bar{t}_{k-1}(J^*)\bar{t}_k(J^*)}{f_{k-1}}\left(\frac{\bar{\beta}_{k-1}(J^*)}{\bar{t}_{k-1}(J^*)} - \frac{\bar{\beta}_k(J^*)}{\bar{t}_k(J^*)}\right)\\
&\leq \Big(a\bar{\beta}_{k-1}(J^*) - \bar{\beta}_k(J^*)\Big)\frac{w}{f^2_k} + \Big(\bar{t}_k(J^*) - \bar{t}_{k-1}(J^*)a\Big)v.
\end{aligned}
$$

Setting $a = \bar{t}_k(J^*)/\bar{t}_{k-1}(J^*)$, we get $f'(0) < 0$. Hence, $\bar{g}(\bar{F}^x) < g_*$ for x small and positive, providing the necessary contradiction. $\square$

Thus, $(\bar{F}^*_{i^*_k} : k \geq 0)$ is a minimizer of the following optimization problem:

$$
\begin{aligned}
\min_{\bar{F}} \quad & \sum_{k=0}^{\infty}\bar{\beta}_k(J^*)/\bar{F}_k \cdot \sum_{k=0}^{\infty}\bar{t}_k(J^*)\bar{F}_k\\
\text{s/t} \quad & \bar{F}_i \geq \bar{F}_{i+1}, \quad \forall i \geq 0\\
& \bar{F}_i > 0, \quad \forall i \geq 0\\
& \bar{F}_0 = 1,
\end{aligned}
$$

This problem is of the same form as (16), except that Proposition 3 now guarantees that $(\bar{\beta}_k(J^*)/\bar{t}_k(J^*) : k \geq 0)$ is strictly decreasing. Proposition 1 then yields the following result.

Theorem 3. *Suppose that $(\bar{\beta}_n : n \geq 0)$ is non-negative and $(\bar{t}_n : n \geq 0)$ is non-decreasing. Then,* *there exists an optimizer* $(\bar{F}_j^* : j \geq 0)$ *to (16) having an associated sequence* J^* *for which*

$$\bar{F}_j^* = \sum_{k=0}^{\infty} \sqrt{\frac{\bar{\beta}_k(J^*)/\bar{t}_k(J^*)}{\bar{\beta}_0(J^*)/\bar{t}_0(J^*)}} \mathbb{I}(i_k^* \leq j < i_{k+1}^*).$$

Furthermore,

$$\bar{g}(\bar{F}^*) = \left(\sum_{k=0}^{\infty} \sqrt{\bar{\beta}_k(J^*)/\bar{t}_k(J^*)} \right)^2$$

Theorem 3 makes clear that the construction of an optimal distribution for N is effectively a "combinatorial problem" that requires finding an optimal sequence J^* minimizing $\sum_{k=0}^{\infty} \sqrt{\bar{\beta}_k(J)/\bar{t}_k(J)}$ over all feasible sequences J for which the $\bar{\beta}_i(J)/\bar{t}_i(J)$'s are decreasing. In practice, the quantities appearing in the formula for $\bar{F}^*$ will need to be estimated from initial "trial run" samples, or by sequentially updating the estimates within an algorithmic implementation in which the distribution of N is constantly readjusted in accordance with the most recent estimators of the $\bar{\beta}_k(J)$'s and $\bar{t}_k(J)$'s.

This leaves open the question of how to efficiently compute the optimal J^*. In a sample setting, only a finite number m of $\bar{\beta}_k$'s and $\bar{t}_k$'s will have been estimated, so we focus exclusively on computing the optimal "m-truncated" sequence $J_m^* = (i_k^* : i_k^* \in \{0, ...m\})$. Algorithm 1 below has a dynamic programming flavor, and recursively computes $J(k,l)$'s and $G(k,l)$'s in the order of increasing k and l values (with $k \leq l$). The quantity $J(k,l)$ stores the best sequence J_l (where J_l is the l-truncation of J) found so far with the last element equal to k, while $G(k,l)$ corresponds to the value of the "cost" $\sum_i \sqrt{\bar{\beta}_i(J(k,l))/\bar{t}_i(J(k,l))}$ associated with $J(k,l)$. If there are no sequences J_l having last element k for which $\bar{\beta}_i(J_l)/\bar{t}_i(J_l)$ is decreasing in i over J_l, we set $J(k,l) = \phi$ (empty sequence) and $G(k,l) = \infty$. It is easily seen that the complexity of this algorithm is of order m^3.

The development of an optimal distribution for N for the independent sum estimator follows a similar path as for the coupled sum estimator, since $\mathbf{var}\,\tilde{Z}$ and $\mathbf{E}\tilde{Z}$ depend on $\mathbf{P}(N \geq \cdot)$ in an exactly similar way.

We conclude this section with a discussion of the optimal distribution for the single term estimator. Here, the associated optimization problem requires finding the optimal probability mass function $(p_n^* : n \geq 0)$ that solves the minimization problem:

$$\begin{aligned} \min_{p} \quad & g(p) \triangleq \left(\sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i^2}{p_i} - \alpha^2 \right) \left(\sum_{i=0}^{\infty} t_i p_i \right) \\ \text{s/t} \quad & p_i > 0, \quad i \geq 0 \\ & \sum_{i=0}^{\infty} p_i = 1. \end{aligned} \tag{19}$$

Theorem 4. *Suppose that* $\mathbf{var}\,\Delta_i > 0$ *for* $i \geq 0$ *and that* $(t_n : n \geq 0)$ *is a positive non-decreasing*

Algorithm 1 Dynamic programming algorithm that finds J_m^* within $O(m^3)$ operations

```
L_{k,l} ← Σ_{j=k}^{l} β̄_j / Σ_{j=k}^{l} t̄_j,   ∀k, l
function OptimalJ
    J(0,0) ← {0}
    G(0,0) ← √(β̄_0 t̄_0)
    for  l = 1 : m do
        J(0,l) ← {0}
        G(0,l) ← √(Σ_{j=0}^{l} β̄_j Σ_{j=0}^{l} t̄_j)
        for  k = 1 : l do
            J(k,l) ← φ
            G(k,l) ← ∞
            P ← √(Σ_{j=k}^{l} β̄_j Σ_{j=k}^{l} t̄_j)
            for  i = 0 : (k − 1) do
                if J(i,k − 1) ≠ φ then
                    J′ ← J(i,k − 1) ∪ {k}
                    G′ ← G(i,k − 1) + P
                    if L_{i,k−1} > L_{k,l} and G′ < G(k,l) then
                        J(k,l) ← J′
                        G(k,l) ← G′
                    end if
                end if
            end for
        end for
    end for
    k* ← arg min_{0≤k≤m} G(k,m)
    return J(k*,m)
end function
```

sequence such that $t_n \to \infty$ *and*

$$\sum_{n=0}^{\infty} \sqrt{\mathbf{E}\Delta_n^2 \cdot t_n} < \infty \tag{20}$$

Then, a minimizer of (19) is the probability mass function $(p_n^* : n \geq 0)$, *where*

$$p_n^* = \sqrt{\frac{\mathbf{E}\Delta_n^2}{\alpha^2 + c^* t_n}} \tag{21}$$

for $n \geq 0$. *Here* c^* *is the unique root of the equation*

$$\sum_{n=0}^{\infty} \sqrt{\frac{\mathbf{E}\Delta_n^2}{\alpha^2 + c^* t_n}} = 1. \tag{22}$$

Furthermore,

$$g(p^*) = c^* \left(\sum_{n=0}^{\infty} t_n p_n^* \right)^2 .$$

Proof. In the presence of (20) and the fact that the t_n's are bounded away from zero, $h(c) \triangleq \sum_{n=0}^{\infty} \sqrt{\mathbf{E}\Delta_n^2/(\alpha^2 + ct_n)}$ is finite-valued and strictly decreasing in $c \geq 0$. Furthermore, the Cauchy-Schwarz inequality implies that

$$\sum_{n=0}^{\infty} \sqrt{\mathbf{E}\Delta_n^2} \;>\; \sum_{n=0}^{\infty} |\mathbf{E}\Delta_n| \;\geq\; |\alpha|,$$

so $h(0) > 1$. Hence, there exists a unique $c^* > 0$ solving (22). Let $\tilde{p} = (\tilde{p}_n : n \geq 0)$ be the probability mass function for which $\tilde{p}_n = \sqrt{\mathbf{E}\Delta_n^2/(\alpha^2 + c^* t_n)}$ for $n \geq 0$, and note that

$$g(\tilde{p}) = \left(\sum_{n=0}^{\infty} \sqrt{\mathbf{E}\Delta_n^2 (\alpha^2 + c^* t_n)} - \alpha^2 \right) \left(\sum_{n=0}^{\infty} \sqrt{\frac{\mathbf{E}\Delta_n^2 \cdot t_n^2}{\alpha^2 + c^* t_n}} \right)$$

Hypothesis (20) implies that $g(\tilde{p}) < \infty$, so that the infimum of $g(\cdot)$ over the feasible region is therefore finite.

As in the proof of Proposition 2, there exists a sequence $(p^{(k)} : k \geq 0)$ of probability mass functions such that $g(p^{(k)})$ converges to the infimum of g over the feasible region. Hence, it follows that there exists $c < \infty$ so that $g(p^{(k)}) \leq c$ for $k \geq k_0 < \infty$. Recall that

$$\mathbf{var}\, Z \;=\; \mathbf{E}[\mathbf{var}\,(Z|N)] + \mathbf{var}\,(\mathbf{E}[Z|N]) \;\geq\; \sum_{n=0}^{\infty} \mathbf{var}\,\Delta_n / p_n \;\geq\; \mathbf{var}\,\Delta_0.$$

So, for $k \geq k_0$,

$$\mathbf{var}\,\Delta_0 \cdot \sum_{n=0}^{\infty} t_n p_n^{(k)} \leq c,$$

so that

$$t_n \sum_{j=n}^{\infty} p_j^{(k)} \leq \frac{c}{\mathbf{var}\,\Delta_0}$$

Since $t_n \to \infty$, the sequence $(p^{(k)} : k \geq 0)$ is tight, so Prohorov's theorem again guarantees the existence of a subsequence $(p^{(n_k)} : k \geq 0)$ and a limiting probability mass function $p^* = (p_j^* : j \geq 0)$ for which $g(p^{(n_k)}) \to g(p^*)$ as $k \to \infty$, and for which p^* attains the infimum of g; clearly, p^* must be feasible.

We now prove that $p^* = \tilde{p}$. For $i \geq 1$, let

$$f_i(q) = \left(\frac{\mathbf{E}\Delta_0^2}{1 - r^* - q} + \sum_{\substack{j=1 \\ j \neq i}} \frac{\mathbf{E}\Delta_j^2}{p_j^*} + \frac{\mathbf{E}\Delta_i^2}{q} - \alpha^2 \right) \cdot \left(t_0(1 - r^* - q) + \sum_{\substack{j=1 \\ j \neq i}}^{\infty} t_j p_j^* + t_i q \right),$$

where $r^* = \sum_{\substack{j=1 \\ j \neq i}}^{\infty} p_j^*$. Clearly, $f_i(p_i^*) = g(p^*)$ and p_i^* is a local minimum of $f_i(\cdot)$. Hence, $f_i'(p_i^*) = 0$. This implies that

$$-\frac{\mathbf{E}\Delta_i^2}{p_i^{*2}} \sum_{j=0}^{\infty} t_j p_j^* + t_i \left(\sum_{j=0}^{\infty} \frac{\mathbf{E}\Delta_j^2}{p_j^*} - \alpha^2 \right) + \frac{\mathbf{E}\Delta_0^2}{p_0^{*2}} \sum_{j=0}^{\infty} t_j p_j^* - t_0 \left(\sum_{i=0}^{\infty} \mathbf{E}\Delta_i^2 - \alpha^2 \right) = 0 \tag{23}$$

for $i \geq 1$. Letting

$$\lambda = \frac{\mathbf{E}\Delta_0^2}{p_0^{*2}} \sum_{j=0}^{\infty} t_j p_j^* - t_0 \left(\sum_{i=0}^{\infty} \mathbf{E}\Delta_i^2 - \alpha^2 \right),$$

multiply the i'th equation in (23) by p_i^* and sum over $i \geq 1$. This yields the equality

$$\lambda \sum_{i=1}^{\infty} p_i^* = \sum_{i=1}^{\infty} \frac{\mathbf{E}\Delta_i^2}{p_i^*} \cdot \sum_{n=0}^{\infty} t_n p_n^* - \sum_{i=1}^{\infty} t_i p_i^* \left(\sum_{n=0}^{\infty} \frac{\mathbf{E}\Delta_n^2}{p_n^*} - \alpha^2 \right). \tag{24}$$

If we add λp_0^* to both sides of (24), we find that

$$\lambda = \alpha^2 \left(\sum_{i=0}^{\infty} t_i p_i^* \right).$$

Plugging into (23), we conclude that p_i^* is given by (21), where $c^* = \left(\sum_{n=0}^{\infty} \frac{\mathbf{E}\Delta_n^2}{p_n^*} - \alpha^2 \right) / \sum_{n=0}^{\infty} t_n p_n^*$. Our expression for $g(p^*)$ then immediately follows. $\square$

As for the summed estimators, in practice, the quantities appearing in the formula for the optimal p* would need to be estimated from either a trial run or within a sequentially updated implementation.

4 Sub-canonical Convergence

As noted in Section 2, our randomized estimators can always be implemented so as to achieve a square-root convergence rate when the strong order p of the approximation error satisfies $p > 1/2$. In fact, when the strong order error is of order p and $\bar{t}_n$ is of order 2^n, Theorem 1 suggests that a good choice for N is to select the distribution so that $\mathbf{P}(N \geq n)$ is order $2^{-n(p+1/2)}$. In particular, for a first order scheme, one would choose N so that $\mathbf{P}(N \geq n) = 2^{-3n/2}$. While this would not be an optimal selection in the sense of Section 3, it can be easily implemented and can be expected to often yield good results. Settings in which the error is of strong order 1 include the Milstein scheme (see p.345 of Kloeden and Platen 1992) and when computing the distribution of the integral of an SDE path for which the underlying SDE can be simulated exactly in discrete time (as for an Asian option based on geometric Brownian motion); see Lapeyre and Temam (2001).

One difficulty with strong p'th order schemes (having $p > 1/2$) is that they can be expensive or impractical to implement. For example, while the Milstein scheme is relatively straightforward to implement when the SDE is driven by a one-dimensional Brownian motion (i.e., $m = 1$), it is challenging to implement when the driving Brownian motion is m-dimensional with $m \geq 2$. In this setting, the approximate time-stepping discretization involves having to generate a collection of iterated Ito integrals of the form

$$\int_0^h B_i(s) dB_j(s),$$

where $B_1, \ldots, B_m$ are m iid standard Brownian motions. Such iterated Ito integrals can be exactly generated when $m = 2$ (see Gaines and Lyons 1994), but no exact algorithm exists when $m > 2$; see Kloeden et al. (1992) and Wiktorsson (2001) for numerically implementable approximations.

As a consequence, we explore in this section the implications of using schemes having a strong order $p \leq 1/2$. This analysis is therefore particularly relevant to SDEs with $m \geq 2$. We focus primarily on understanding the computational complexity improvements that can be obtained by applying randomization methods in the SDE setting. For this purpose, we exclusively analyze the single-term estimator, because it illustrates the main points and the resulting calculations are easier to carry out. However, in our SDE context, one would expect similar complexity results to hold for the two summed estimators.

As noted in Section 2, it is always straightforward in the SDE setting to construct finite variance unbiased estimators for α, once p is known (Just choose $\mathbf{P}(N \geq n) = 2^{-nr}$, with $r > 2p$). So, reliable confidence interval methodologies and sequential stopping procedures (for achieving a given error tolerance ϵ) can always be implemented in the SDE setting. The key remaining question is the design of randomized algorithms that can achieve a low complexity. Chebyshev's inequality implies that in order to achieve ϵ error with probability $1-\delta$ (with $0<\delta<1$) (for given ϵ and δ fixed), one must choose the number n of iid samples of Z of order $1/\epsilon^2$. The question is: how much "work" must be done in order to generate $n = O(1/\epsilon^2)$ samples? Furthermore, how can we design randomized algorithms that will minimize this work complexity while maintaining finite variance?

We start by discussing the design of the distribution of N for the Euler discretization scheme; see p.340 of Kloeden and Platen (1992). This scheme is the simplest of all SDE schemes to implement, and perhaps the most widely applied. This has an associated strong order $p = 1/2$, regardless of the value of m. We take the view here that the work per Euler approximation equals the number of time steps, so that the time t_n needed to generate Δ_n is $2^{n-1} + 2^n$. To be precise regarding the variance, we now specialize to the case where the functional f of Section 1 is suitably Lipschitz. In this case, $\|\Delta_n\|_2 = O(2^{-n/2})$; see Ben Alaya and Kebaier (2012). We need to choose the probability mass function of N so that the total work $W_n \triangleq t_{N_1} + ... + t_{N_n}$ needed to generate $n = O(1/\epsilon^2)$ iid copies of Z grows as slowly as possible as $\epsilon \to 0$, while maintaining finite variance for Z. We set p_i in proportion to $2^{-i}i(\log_2(1+i))^2$, and note that $\mathbf{var}\, Z$ is then finite. It is easy to see that $\mathbf{P}(N \geq i)$ is of order p_i as $i \to \infty$.

To study the rate of growth of W_n, we apply the following result due to Feller (1946).

Result 1 (Result.)**.** Suppose that $(a_n : n \geq 0)$ is a sequence for which a_n/n increases as $n \to \infty$. Then, $W_n \leq a_n$ eventually w.p. 1—i.e., $\mathbf{P}(W_n > a_n \text{ infinitely often}) = 0$—if $\sum_{n=0}^{\infty} \mathbf{P}(t_N \geq a_n) < \infty$.

Note that the conclusion implies $\mathbf{P}(W_n > a_n) \to 0$ as $n \to \infty$. In view of this, we declare that the complexity is $O(a_n)$ if we can find a_n such that the infinite sum in the Result is finite.

Since $t_n = 2^n + 2^{n-1}$, $\mathbf{P}(t_N \geq x) = \Theta(x^{-1}\log_2 x(\log_2\log_2 x)^2)$ where $f(n) = \Theta(g(n))$ means that $g(n)$ is bounded from above and below by constant multiple of $f(n)$ for sufficiently large n's (i.e, there exist constants c_1, c_2, and N_0 such that $c_1 f(n) \leq g(n) \leq c_2 f(n)$ for $n \geq N_0$), and hence, if we choose $a_n = n(\log_2 n)^q$ with $q > 2$, it is easily seen that $\mathbf{P}(t_N \geq a_n)$ is a summable sequence, yielding the following result.

Proposition 4. *Fix $q > 2$. When $\|\Delta_n\|_2 = O(2^{-n/2})$, a single-term estimator can be defined for which the computational complexity required to compute $\mathbf{E}Y$ to within ϵ with probability $1-\delta$ is $O\big((1/\epsilon^2)(\log_2(1/\epsilon))^q\big)$ as $\epsilon \to 0$.*

We turn next to the complexity estimate for the single-term estimator that can be achieved when $\|\Delta_n\|_2 = O(2^{-np})$ for $p \in (0, 1/2)$. Such strong orders arise, for example, when computing the value of a digital option in which the underlying SDE is approximated via the Euler scheme; see Giles (2008b). We further assume here that the bias of Y_n is such that $\mathbf{E}(Y_n - Y) = O(2^{-ns})$ for $s \geq 1/2$. Again, we take the view that the work required to compute Y_n is of order 2^n. In this setting, we introduce a new tactic that can be used to shape the design of the associated single-term estimator, so as to minimize the complexity: conditional on $N = k$, we can draw multiple iid replicates from the population of Δ_k, average them, and return the average divided by p_k (where the sequence $(m_k : k \geq 0)$ is carefully chosen). More specifically, we apply sample size $m_k = 2^{\lfloor \gamma k \rfloor}$ to the k'th element of the difference sequence. As a consequence, $\|\bar{Y}_k - \bar{Y}_{k-1}\|_2^2 = O(2^{-(2p+\gamma)k})$ (where $\bar{Y}_k - \bar{Y}_{k-1}$ is a sample mean of m_k iid replicates of $Y_k - Y_{k-1}$), provided that $(\mathbf{E}(\bar{Y}_k - \bar{Y}_{k-1}))^2$ is of the same order (or smaller) than the variance; this occurs precisely when $\gamma + 2p \leq 2s$, so that the variance of $\bar{Y}_k - \bar{Y}_{k-1}$ is the dominant contribution to the L^2 norm of $\bar{Y}_k - \bar{Y}_{k-1}$. On the other hand, when $\gamma + 2p > 2s$, then $\|\bar{Y}_k - \bar{Y}_{k-1}\|_2^2 = O(2^{-2sk})$. It is evident that increasing the sample size to a point at which the variance of $\bar{Y}_k - \bar{Y}_{k-1}$ is much smaller than its squared mean is a waste of computational effort, so we should constrain γ so that $\gamma + 2p \leq 2s$.

With this choice of sample size and subsequence, the work t_N per sample mean computed is of order $2^{(\gamma+1)N}$. Analogously to the discussion of the case in which $p = 1/2$, we set p_i in proportion to $2^{-(2p+\gamma)i} i(\log_2(1+i))^2$. If $a_n = n^v (\log_2 n)^w$ for $v, w > 0$, then $\mathbf{P}(t_N \geq a_n) = \mathbf{P}(N \geq \frac{v}{\gamma+1}\log_2 n + \frac{w}{\gamma+1}\log_2\log_2 n + O(1))$ for $O(1)$ a deterministic function of n. This is asymptotically

$$n^{-v(2p+\gamma)/(\gamma+1)}(\log_2 n)^{-w(2p+\gamma)/(\gamma+1)+1}(\log_2\log_2 n)^2 \tag{25}$$

as $n \to \infty$. The sequence $\mathbf{P}(t_N > a_n)$ is summable if we choose $v = (\gamma+1)/(2p+\gamma)$ and w so that $-w(2p+\gamma)/(\gamma+1) + 1 < -1$. The exponent v that fundamentally determines the growth rate of a_n can be made smallest by choosing $\gamma = 2(s-p)$, in which case $v = 1 + (1-2p)/2s$. Given the above constraint on w and the fact that r can be made arbitrarily large, w must be set greater than $2(\gamma+1)/(2p+\gamma) = 2 + (1-2p)/s$. Another application of the result above due to Feller yields our second complexity result.

Proposition 5. *Fix $a > 0$. When $\|\Delta_n\|_2 = O(2^{-np})$ for $p \in (0, 1/2)$ and $\mathbf{E}(Y_n - Y) = O(2^{-ns})$ for $s \geq 1/2$, a single-term estimator can be defined for which the computational complexity required to compute $\mathbf{E}Y$ to within ϵ with probability $1-\delta$ is $O\big((1/\epsilon)^{2+(1-2p)/s}(\log_2(1/\epsilon))^{2+(1-2p)/s+a}\big)$ as $\epsilon \to 0$.*

The complexity results in Proposition 4 and 5 are very close to that obtained by Giles (2008b) for MLMC in similar settings; our complexity results contain the extra logarithmic factors. The discussion of this section therefore suggests that the complexity theory for our randomization methods looks very similar to that which has been obtained for MLMC.

5 SDE Implementation and Computational Results

As noted earlier, our randomization methods take advantage of pairwise couplings between Y_i and Y_{i-1} in order to obtain an unbiased estimator. In the SDE setting, Y_i is most naturally constructed from a time-discretization involving increments of length 2^{-i}. The Y_i's therefore involve different levels of refinement with regard to the time discretization. This theme, in which one runs simulations at different levels of refinement, is also central to the construction of MLMC algorithms.

In some sense, our methods can be viewed as randomized versions of MLMC algorithms, in which the number of levels used by MLMC is randomly determined. While MLMC constructs biased estimators with a carefully controlled (and optimized) level of bias, our approach is to construct unbiased estimators to which the full theory of conventional Monte Carlo can be applied. For example, in Section 2, we described how asymptotically valid (fully rigorous) confidence interval procedures can be easily developed for our estimators, based on either fixed sample size or sequential settings (designed to achieve a given level of either absolute or relative precision); developing analogous procedures in the presence of bias (that is of the same order as the variability) is more complex. In contrast to MLMC, the estimators developed here are independent of the level ϵ of precision needed, while those associated with MLMC are constructed relative to a given ϵ tolerance. Such MLMC algorithms therefore require more analysis in order to carefully assess the bias, and to control for it. (In addition, when a practical MLMC implementation uses sample-based estimates to calibrate the parameters needed to achieve a given ϵ error, the resulting estimator no longer is covered by the theoretical guarantees associated with the analyses that assume a priori knowledge of various problem parameters; see Giles 2008a) Furthermore, our approach leads naturally to statistical formulations within which optimal design choices (e.g. the optimal distributions for N studied in Section 3) can be made.

Nevertheless, as suggested by the complexity analysis of Section 4, the performance of the randomization methods studied here can be expected to share many of the theoretical and empirical properties of MLMC. In particular, the couplings that have been successfully applied within the MLMC setting can be expected to be equally valuable in our context. The choice of the appropriate coupling to be used depends crucially on both the number m of independent driving Brownian motions and the form of the path functional f mentioned in Section 1. We say that f is of "final value" form if $f(x) = v(x(1))$ for some smooth given real-valued (deterministic) function v, while f is of "integral form" if $f(x) = v(\int_0^1 w(x(s))ds)$ for v, w deterministic. In the particular case that $v(y) = [y_1 - k]^+$ for some positive constant k (where y_1 is the first component of y), we refer to such a final value functional as a European option. When $v(y) = [y - k]^+$ and $w(y) = y_1$, we call such an integral functional an Asian option. Three additional functionals are also widely used within the computational finance community, specifically lookback options (i.e., $f(x) = x_1(1) - \min\{x_1(s) : 0 \le s \le 1\}$), digital options (i.e., $f(x) = \mathbb{I}(x(1) \in B)$ for some given subset B), and barrier options (i.e., $f(x) = [x_1(1) - k]^+ \mathbb{I}(\tau(x) > 1)$ also, $\tau(x) = \inf\{t \ge 0 : x(t) \in B\}$ for some given subset B).

When f is a final value functional for which v is Lipschitz and and $m \le 2$, one can apply the standard Milstein scheme to obtain an appropriate sequence of Y_i's satisfying the conditions of Section 2. This requires a probability space that simultaneously supports both Y and all the Y_i's . While the standard proofs that the Milstein scheme achieves strong order 1 only consider the joint distribution of (Y_i, Y), a perusal of the argument makes clear that one could equally well have constructed a single probability space supporting B, Y, and Y_1, Y_2, ... under which $\|Y_n - Y\|_2 = O(2^{-n})$ (thereby implying that $\|\Delta_n\|_2 = O(2^{-n})$); see p.363 of Kloeden and Platen (1992) for example. As noted in Section 4, when $m > 2$, then the conventional Milstein method becomes difficult to apply directly, because of the presence of the iterated Ito integrals that must be generated. Fortunately, in this setting, a newly proposed antithetic truncated Milstein scheme due to Giles and Szpruch (2012) is potentially applicable when v is both Lipschitz and (appropriately) smooth. It is strong first order in this context, and is essentially order 3/4 when v is not smooth. However, it should be noted that this scheme does not directly fall into the framework of Section 2, because the approximating rvs are not L^2 approximations of the rv Y under consideration. Rather, one needs to recognize that the key elements in deriving the three estimators discussed in Section 2 fundamentally hinge upon only two facts. The first is the existence of a sequence of rvs Y_n for which $\mathbf{E}Y_n$ converges to $\mathbf{E}Y$ as $n \to \infty$; the second is the need for a sequence of rvs $(\Delta'_n : n \geq 0)$ for which $\mathbf{E}\Delta'_n = \mathbf{E}(Y_n - Y_{n-1})$ with $\|\Delta'_n\|_2 \to 0$ sufficiently quickly. We can then generalize upon Section 2's randomization methods by substituting the Δ'_n's for the Δ_n's in Section 2's estimators. (We chose to not introduce the theory in Section 2 at this more general level, in order to ease the exposition.)

We now generalize the discussion of Section 2 to cover this modified setting. As in Section 2, we assume that N is independent of the sequence $(\Delta'_i : i \geq 0)$. To maximize the potential applicability of this result (to settings outside the SDE context), we do not require here that the object α to be computed be expressible as the expectation $\mathbf{E}Y$ of some rv Y. Rather, we permit α here to be a quantity that can be expressed as a limit of the expectations $\mathbf{E}Y_n$. (For example, the density of a rv at a given point can be expressed as such a limit without being expressible in the form $\mathbf{E}Y$.)

Theorem 5. *Assume that Y_n is integrable for each n and suppose that $(\Delta'_i : i \geq 0)$ is a sequence of rvs for which $\mathbf{E}\Delta'_i = \mathbf{E}Y_i - \mathbf{E}Y_{i-1}$ for $i \geq 0$.*

a) If $(\Delta'_i : i \geq 0)$ is a sequence of independent rvs for which there exists α such that

$$\sum_{i=0}^{\infty}(\|\Delta'_i\|_2^2 + (\mathbf{E}Y_i - \alpha)^2)/\mathbf{P}(N > i) < \infty, \tag{26}$$

then $\mathbf{E}Y_i$ converges to α as $i \to \infty$, $\tilde{Z}' \triangleq \sum_{i=0}^{N} \Delta'_i/\mathbf{P}(N \geq i)$ is an unbiased estimator for α, and

$$\mathbf{E}(\tilde{Z}')^2 = \sum_{i=0}^{\infty} \tilde{\nu}'_i/\mathbf{P}(N \geq i),$$

where $\tilde{\nu}'_i = \mathbf{var}\,\Delta'_i + (\alpha - \mathbf{E}Y_{i-1})^2 - (\alpha - \mathbf{E}Y_i)^2$.

b) If $(\sum_{i=0}^{n} \Delta'_i : n \geq 0)$ is a Cauchy sequence in L^2 converging to a limit Y' (say) that further *satisfies*

$$\sum_{i=0}^{\infty} \frac{\| \sum_{j=i+1}^{\infty} \Delta_j' \|_2^2}{\mathbf{P}(N > i)} < \infty, \tag{27}$$

then $\bar{Z}' \triangleq \sum_{i=0}^{N} \Delta_i'/\mathbf{P}(N \geq i)$ *is an unbiased estimator for* $\alpha \triangleq \lim_{n\to\infty} \mathbf{E} \sum_{i=0}^{n} \Delta_i'$ *and*

$$\mathbf{E}(\bar{Z}')^2 = \sum_{i=0}^{\infty} \bar{\nu}_i'/\mathbf{P}(N \geq i), \tag{28}$$

where $\bar{\nu}_i' \triangleq \|Y_{i-1}' - Y'\|_2^2 - \|Y_i' - Y'\|_2^2$ *and* $Y_i' \triangleq \sum_{j=0}^{i} \Delta_j'$.

Alternatively, if a sequence $(\sum_{i=0}^{n} \Delta_i' : n \geq 0)$ *satisfies*

$$\sum_{i=0}^{\infty} \|\Delta_i'\|_2^2/\mathbf{P}(N \geq i) + \sum_{i<j} \|\Delta_i'\|_2 \|\Delta_j'\|_2/\mathbf{P}(N \geq i) < \infty, \tag{29}$$

then $Y_i' \triangleq \sum_{j=0}^{i} \Delta_j'$ *converges to a limit* Y' *in* L^2, *and* $\bar{Z}' \triangleq \sum_{i=0}^{N} \Delta_i'/\mathbf{P}(N \geq i)$ *is an unbiased estimator for* α *with second moment (28).*

c) If $\sum_{i=0}^{\infty} \mathbf{E}(\Delta_i')^2/\mathbf{P}(N = i) < \infty$, *then there exists* $\alpha \in \mathbb{R}$ *for which* $\mathbf{E}Y_i \to \alpha$ *as* $i \to \infty$, $Z' \triangleq \Delta_N'/p_N$ *is an unbiased estimator for* α *and*

$$\mathbf{E}Z'^2 = \sum_{i=0}^{\infty} \mathbf{E}(\Delta_i')^2/\mathbf{P}(N = i).$$

Proof. The proof of part a is similar to that of Theorem 2, except that here the hypotheses are stated in terms of the Δ_i''s (rather than the δ_i's used there). Put $b_i = \mathbf{E}Y_i - \alpha$ and note that (26) implies that $\mathbf{E}Y_i \to \alpha$ as $i \to \infty$. Put $\rho_0 = 0$ and $\rho_k = \inf\{j > \rho_{k-1} : |b_j| \leq |b_{\rho_{k-1}}|\}$ for $k \geq 1$. The proof of Theorem 2 shows that for $n > m$,

$$
\begin{aligned}
&\left\| \sum_{i=0}^{\rho_n} \Delta_i' \mathbb{I}(N \geq i)/\mathbf{P}(N \geq i) - \sum_{i=0}^{\rho_m} \Delta_i' \mathbb{I}(N \geq i)/\mathbf{P}(N \geq i) \right\|_2^2 \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[(\Delta_i')^2 + 2\mathbf{E}\Delta_i'(\mathbf{E}Y_{\rho_n} - \mathbf{E}Y_i)]/\mathbf{P}(N \geq i) \\
&= \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[(\Delta_i')^2 + 2(b_i - b_{i-1})(b_{\rho_n} - b_i)]/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[(\Delta_i')^2 + (2b_{\rho_n}^2 + 2b_{i-1}^2)]/\mathbf{P}(N \geq i) \\
&\leq \sum_{i=\rho_m+1}^{\rho_n} \mathbf{E}[(\Delta_i')^2 + 4b_{i-1}^2]/P(N \geq i),
\end{aligned}
$$

where the first inequality follows from the fact that $2ab \leq a^2 + b^2$ for $a, b \in \mathbb{R}$, and the second is a consequence of the definition of the ρ_n's. In view of (26), $(\sum_{i=0}^{\rho_n} \Delta_i'/\mathbf{P}(N \geq i) : n \geq 0)$ is Cauchy in L^2. From this point onwards, the proof is identical to that of Theorem 2.

Turning now to the first statement of part b, the fact that the Y_i''s are Cauchy implies that Y_i' converges to Y' as $i \to \infty$, and that $\mathbf{E}Y_i'(= \mathbf{E}Y_i)$ converges to a a limit α as $i \to \infty$. We now let Y' and $(Y_i' : i \geq 0)$ play the role of Y and $(Y_i : i \geq 0)$, respectively, in Theorem 1, and apply Theorem 1 to prove the statement. For the second statement of part b, note that the alternative condition (29) implies $\sum_{i=0}^{\infty} \|\Delta_i'\|_2 < \infty$ and hence that Y_i''s are Cauchy. Now, a similar (but simpler) argument as the one for Theorem 1 finishes the proof of b. For part c, it is easy to see that $\mathbf{E}(\Delta_N'/p_N)^2 = \sum_{i=0}^{\infty} \mathbf{E}(\Delta_i')^2/\mathbf{P}(N = i)$. The hypothesis therefore guarantees that $\mathbf{E}(\Delta_N'/p_N)^2 < \infty$, so that $\mathbf{E}|\Delta_N'|/p_N < \infty$. It follows that $\sum_{i=0}^{\infty} |\mathbf{E}\Delta_i'| < \infty$, so that $\mathbf{E}Y_i = \sum_{j=0}^{i} \mathbf{E}\Delta_j'$ converges to a limit α. The rest of part c follows easily. $\square$

The specific form of the rv Δ_i' that arises in the setting of the antithetic truncated Milstein estimator is $\frac{1}{2}\big(f(X_{h_i}) - f(\tilde{X}_{h_i})\big) - f(X_{h_{i-1}})$, where h_i is the time-step used for the truncated Milstein scheme at level i and $\tilde{X}_{h_i}$ is the antithetic version of X_{h_i} obtained (conditional on $X_{h_{i-1}}$) by using the finer Brownian increments needed at level h_i in reverse order relative to the finer Brownian increments used by X_{h_i}. When Theorem 4.10 of Giles and Szpruch (2013) applies, it follows that when $h_i = 2^{-i}$, $\|\Delta_i'\|_2$ is of order 2^{-i}. This implies that $\mathbf{E}|\Delta_i'|$ is also of order 2^{-i}, and hence all three estimators described in Theorem 5 are then applicable, provided that the distribution of N is appropriately selected. In particular, one can easily select N so that all three estimators then enjoy square root convergence rates. Finally, because the expressions for the variances of the three estimators of Theorem 5 are identical to those of Section 2, the theory of Section 3 on optimally selecting the distribution of N is also applicable in this context.

The above discussion has been focused on final value expectations. In computing expectations of more general path functionals, it should be noted that the the quality of the Euler and Milstein schemes do not degrade when looking at the quality of the approximation across the entire set of discretization points, in the sense that $\| \max\{|X_h(ih) - X(ih) : 0 \le ih \le 1\}\|_2 = O(\|X_h(1) - X(1)\|_2)$ as $h \to 0$; see Theorem 10.6.3 of Kloeden and Platen (1992). The challenge with Asian, lookback, digital, and barrier options is that the SDE path behavior between discretization points introduces an error of order $h^{1/2}$, which immediately leads to a strong order $p = 1/2$, regardless of whether a higher order scheme has generated the approximating path at the discretization points or not. Thus, one needs to generate some additional approximating rvs within each subinterval $[ih, (i+1)h]$ to capture the principal path fluctuation effects for that subinterval associated with the specific path functional under consideration. These additional approximating rvs are described in Giles (2008a) for each of these four options that depend on SDE path behavior between discretization epochs.

Having discussed how our theory specifically applies in the SDE setting, we now report on our computational experience with this class of methods. We implemented each of our three estimators, as well as the MLMC method, on the following SDE models, all of which are widely used in finance:

Example 1. *(Geometric Brownian motion (gBM)): Here, the SDE for X is*

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

with the parameters selected as $\mu = 0.05$, $\sigma = 0.2$, and $X(0) = 1$. Our focus here is on computations for the final value "European option" $f(x) = \exp(-\mu)[x(1) - 1]^+$, having computed value $\mathbf{E}f(X) = 0.104505836$. While X_h in this setting can be exactly simulated, we have instead applied the standard Milstein scheme.

Example 2. *(Cox-Ingersoll-Ross (CIR) process): The SDE for X is in this case given by*

$$dX(t) = \kappa(\theta - X(t))dt + \sigma X(t)^{1/2} dB(t),$$

with parameters given by $\mu = 0.05$, $\kappa = 5$, $\theta = 0.04$, $\sigma = 0.25$, and $X(0) = 0.04$. We provide here numerical results for the European option $f(x) = \exp(-\mu)[x(1) - 0.03]^+$; the standard Milstein scheme underlies our discretization X_h. The quantity $\mathbf{E}f(X) = 0.0120124$ was computed using the coupled-sum estimator with a target root mean square error (RMSE) 1.0×10^{-7}.

Example 3. *(Heston model): This is a two-dimensional model for which*

$$\begin{aligned} dS(t) &= \mu S(t) + V(t)^{1/2} S(t) dB_1(t) \\ dV(t) &= \kappa(\theta - V(t))dt + \sigma V(t)^{1/2} dB_2(t), \end{aligned}$$

where B_1 and B_2 are correlated Brownian motions with correlation $\rho = -0.5$. The specific parameter values used here are $\mu = 0.05$, $\kappa = 5$, $\theta = 0.04$, $\sigma = 0.25$, $S(0) = 1$, $V(0) = 0.04$, and we apply this to the specific functional $f(s, v) = \exp(-\mu)[s(1) - 1]^+$. For this example, we utilized the antithetic truncated Milstein scheme mentioned earlier to generate the Δ_i''s. The value of $\mathbf{E}f(S, V)$ is 0.10459672; see Kahl and Jäckel (2006).

In order to make the computational comparison with MLMC as transparent as possible, we adopt the approach commonly followed within the MLMC literature, in which the parameters of MLMC algorithm are set so that the resulting estimator will possess a specific root mean square error (RMSE) ϵ; see Giles (2008b) for details. Of course, as mentioned earlier, the way in which the parameters are set depends upon unknown model-specific quantities that are estimated online and used to adaptively modify the parameters within the algorithm. Consequently, the final RMSE for MLMC may differ (significantly) from the intended RMSE. Turning to our three randomization algorithms, we noted in Section 2 that these methods are well suited to sequential implementations in which the algorithms are run until the $100(1-\delta)\%$ confidence interval half-width is less than or equal to η. The resulting estimator, denoted (for example) $\bar{\alpha}_{K(\eta)}$ when applied to the coupled-sum algorithm, is not intended to produce an estimator with a given level of RMSE. However, when η is chosen as $z\epsilon$ (with z satisfying $\mathbf{P}(-z \le N(0,1) \le z) = 1 - \delta$), this corresponds to sampling until the estimated standard deviation is less than or equal to ϵ. In our numerical comparisons, we set η in this way for our sequential implementations of our three estimators, in order to simplify the comparison with MLMC. The confidence level for all the confidence interval methodologies discussed in this section is set at 90%. In addition, in order to prevent our sequential stopping procedure from terminating early because of an unreliable small-sample estimate of the standard deviation, we modify $K(\eta)$ so that at least 1000 iid samples are generated before we begin testing to see if the half-widths are less than η or not.

To determine the optimal distribution of N for the coupled-sum estimator, we have estimated the $\bar{\nu}_i$'s for the first few i's with 10,000 samples and extrapolated using asymptotic rates. Specifically, since Y can not be sampled exactly, we have used $\bar{\nu}_i^\dagger = \|Y_{13} - Y_{i-1}\|^2 - \|Y_{13} - Y_i\|^2$ as a surrogate for $\bar{\nu}_i = \|Y - Y_{i-1}\|^2 - \|Y - Y_i\|^2$ for $i = 0, 1, \cdots, 8$. For $i > 8$, we have extrapolated the $\bar{\nu}$ sequence using the approximation $\bar{\nu}_{8+j} \approx \bar{\nu}_8 \times 2^{-2jp}$ where p is the strong order. To determine N for the independent-sum estimator, we have computed the $\tilde{\nu}_i$'s by estimating $\mathbf{var}\,\tilde{\Delta}_i$ and $\mathbf{E}\tilde{\Delta}_i$ from 10,000 samples for $i = 0, \cdots, 10$. Then, we have extrapolated again using the approximations $\mathbf{var}\,\tilde{\Delta}_{10+j} \approx \mathbf{var}\,\tilde{\Delta}_{10} \times 2^{-2jp}$ and $\mathbf{E}\Delta_{10+j} \approx \mathbf{E}\Delta_{10} \times 2^{-js}$ for $j \ge 1$ where p is the strong order and s is the weak order. Once the parameters were estimated, we have used Algorithm 1 (with $m = 10$) with these parameters as input to find the optimal distribution of N for summed estimators. For the single-term estimator, the first ten $\mathbf{E}\Delta_i^2$'s were estimated with 10,000 samples in an obvious way, and extrapolated using the approximation $\mathbf{E}\Delta_{10+j}^2 \approx \mathbf{E}\Delta_{10}^2 \times 2^{-2jp}$ with $p = 1$ for $j \ge 1$. After the parameters were estimated, we used Newton's method to find c^* in (21).

The computational results are reported in Tables 1-4 for gBM, Tables 5-8 for CIR, and Tables 9-12 for Heston. For each of our problems, the discretization used at level i (to generate Y_i) involved time step 2^{-i}. In each table, the first column represents the quantity q, in which the corresponding row relates to calculations intended to generate a RMSE of approximately magnitude $q|\alpha|$; we use IRE for this column as an abbreviation for "intended RMSE". Thus, our sequential algorithms are run with $\eta = zq|\alpha|$, while MLMC is designed to achieve RMSE $\epsilon = q|\alpha|$. Specifically, we have implemented MLMC based on Giles' MATLAB implementation available at `http://people.maths.ox.ac.uk/gilesm/files/mcqmc06_code.zip`. For each of the four algorithms, the second column represents a confidence interval for the expectation of the computed solution, based on 1000 iid replications of the algorithm based on the IRE level specified. The third column is the sample RMSE, divided by $|\alpha|$, estimated from the 1000 replications (with deviations measured relative to the true solution), while the fourth column is the sample standard deviation, divided by $|\alpha|$, as computed from the 1000 replications. Column 5 is relevant only to the MLMC tables, and provides the estimated bias divided by $|\alpha|$, as determined by the average of the 1000 replications and the exact solution. The sixth column is a confidence interval for the expected work expended, based on the 1000 replications, at a given IRE level; we took, as a measure of the work expended in a given replication, the total number of time steps simulated for that replication. The final column, denoted Work$\times$MSE, is the product of the estimated expected work per replication and the estimated mean square error (MSE). For an algorithm exhibiting a square root convergence rate, this product should be asymptotically constant, as the level q shrinks to 0.

Further computational results can be found in the online supplement. The online supplement provides additional computational results for one more model (the Vasicek model), and for additional path functionals beyond the final value functionals described in this paper (specifically, results related to Asian, lookback, digital, and barrier options). The computational experiments confirm that our algorithms do indeed produce unbiased estimators and achieve square root convergence. Moreover, the experiments suggest that our estimators are competitive with multilevel Monte Carlo methods, at least for the examples studied, in terms of the work-MSE achieved. To be precise, the work-MSE factors are roughly identical, except for the gBM example in which the work-MSE factor for all three randomization methods is about 60–70% that of MLMC, while the coupled-sum estimator is about ten times more efficient than MLMC for the CIR example (while the other two randomization methods are roughly equivalent to MLMC). It is also worth noting that MLMC often "overshoots" the desired IRE, while the sample-based sequential stopping criterion used for our randomization algorithms more nearly matches the IRE. Thus, for a given desired accuracy, our unbiased estimators lend themselves to implementations that meet the desired error tolerance, without doing additional computation that will refine the accuracy beyond that needed. Given that these unbiased randomization methods lend themselves readily to the incorporation of the full spectrum of output analysis and variance reduction techniques that are standard tools in the setting of conventional iid Monte Carlo algorithms, the use of randomization in the SDE setting as introduced in this paper seems a promising direction for future research.

Acknowledgments

The authors gratefully acknowledge the support of National Science Foundation Grant DMS-1320158 and a Samsung Scholarship.

References

Ben Alaya, M., A. Kebaier. 2012. Central limit theorem for the multilevel Monte Carlo Euler method and applications to Asian options. Submitted for publication.

Beskos, A., G. Roberts. 2005. Exact simulation of diffusions. *Annals of Applied Probability* **15**(4) 2422–2444.

Billingsley, P. 1999. *Convergence of Probability Measures*. 2nd ed. John Wiley and Sons, New York.

Chen, N., Z. Huang. 2013. Localization and exact simulation of Brownian motion driven stochastic differential equations. *Mathematics of Operations Research* Forthcoming.

Duffie, D., P. W. Glynn. 1995. Efficient Monte Carlo simulation of security prices. *Annals of Applied Probability* **5**(4) 897–905.

Feller, W. 1946. A limit theorem for random variables with infinite moments. *American Journal of Mathematics* **68**(2) 257–262.

Gaines, J. G., T. J. Lyons. 1994. Random generation of stochastic integrals. *SIAM Journal of Applied Mathematics* **54**(4) 1132–1146.

Giesecke, K., D. Smelov. 2013. Exact sampling of jump-diffusions. *Operations Research* Forthcoming.

Giles, M. B. 2008a. Improved multilevel Monte Carlo convergence using the Milstein scheme. A. Keller, S. Heinrich, H. Niederreiter, eds., *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Springer-Verlag, 343–358.

Giles, M. B. 2008b. Multilevel Monte Carlo path simulation. *Operations Research* **56**(3) 607–617.

Giles, M. B., L. Szpruch. 2012. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. Submitted for publication.

Glynn, P. W. 1983. Randomized estimators for time integrals. Tech. rep., Mathematics Research Center, University of Wisconsin, Madison.

Glynn, P. W., W. Whitt. 1992a. The asymptotic efficiency of simulation estimators. *Operations Research* **40** 505–520.

Glynn, P. W., W. Whitt. 1992b. The asymptotic validity of sequential stopping rules for stochastic simulations. *Annals of Applied Probability* **2** 180–198.

Kahl, C., P. Jäckel. 2006. Fast strong approximation Monte-Carlo schemes for stochastic volatility models. *Quantitative Finance* **6**(6) 513–536.

Kloeden, P. E., E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer-Verlag.

Table 1: gBM, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.10433 $\pm 2.4\times10^{-4}$	4.5×10^{-2}	4.5×10^{-2}	-	$1.5\times10^{3} \pm 4.2\times10^{1}$	0.032
0.0200	0.10427 $\pm 1.0\times10^{-4}$	1.9×10^{-2}	1.9×10^{-2}	-	$9.5\times10^{3} \pm 1.6\times10^{2}$	0.038
0.0100	0.104488 $\pm 5.1\times10^{-5}$	9.3×10^{-3}	9.4×10^{-3}	-	$3.6\times10^{4} \pm 4.2\times10^{2}$	0.034
0.0050	0.104496 $\pm 2.3\times10^{-5}$	4.3×10^{-3}	4.3×10^{-3}	-	$1.6\times10^{5} \pm 1.2\times10^{3}$	0.033
0.0020	0.104505 $\pm 1.1\times10^{-5}$	2.0×10^{-3}	2.0×10^{-3}	-	$8.6\times10^{5} \pm 6.3\times10^{3}$	0.036
0.0010	0.1045083 $\pm 4.6\times10^{-6}$	8.5×10^{-4}	8.5×10^{-4}	-	$4.2\times10^{6} \pm 5.7\times10^{3}$	0.033
0.0005	0.1045082 $\pm 2.7\times10^{-6}$	5.0×10^{-4}	5.0×10^{-4}	-	$1.3\times10^{7} \pm 3.6\times10^{4}$	0.035

Table 2: gBM, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.10462 $\pm 2.3\times10^{-4}$	4.3×10^{-2}	4.3×10^{-2}	-	$1.4\times10^{3} \pm 2.0\times10^{1}$	0.029
0.0200	0.10437 $\pm 1.0\times10^{-4}$	1.9×10^{-2}	1.9×10^{-2}	-	$8.5\times10^{3} \pm 1.1\times10^{2}$	0.032
0.0100	0.104478 $\pm 4.9\times10^{-5}$	9.1×10^{-3}	9.1×10^{-3}	-	$3.5\times10^{4} \pm 2.0\times10^{2}$	0.032
0.0050	0.104485 $\pm 2.6\times10^{-5}$	4.8×10^{-3}	4.8×10^{-3}	-	$1.2\times10^{5} \pm 1.4\times10^{3}$	0.031
0.0020	0.1045056 $\pm 9.7\times10^{-6}$	1.8×10^{-3}	1.8×10^{-3}	-	$9.0\times10^{5} \pm 2.9\times10^{3}$	0.031
0.0010	0.1045041 $\pm 5.4\times10^{-6}$	9.9×10^{-4}	9.9×10^{-4}	-	$3.0\times10^{6} \pm 1.0\times10^{4}$	0.033
0.0005	0.1045073 $\pm 2.6\times10^{-6}$	4.9×10^{-4}	4.9×10^{-4}	-	$1.2\times10^{7} \pm 1.1\times10^{4}$	0.031

Table 3: gBM, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.10450 $\pm 2.3\times10^{-4}$	4.3×10^{-2}	4.3×10^{-2}	-	$1.3\times10^{3} \pm 1.4\times10^{1}$	0.026
0.0200	0.104373 $\pm 9.9\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$8.2\times10^{3} \pm 1.1\times10^{2}$	0.030
0.0100	0.104470 $\pm 5.0\times10^{-5}$	9.2×10^{-3}	9.2×10^{-3}	-	$3.2\times10^{4} \pm 2.0\times10^{2}$	0.029
0.0050	0.104493 $\pm 2.7\times10^{-5}$	4.9×10^{-3}	4.9×10^{-3}	-	$1.1\times10^{5} \pm 1.1\times10^{3}$	0.029
0.0020	0.1044989 $\pm 9.6\times10^{-6}$	1.8×10^{-3}	1.8×10^{-3}	-	$8.1\times10^{5} \pm 1.4\times10^{3}$	0.028
0.0010	0.1045129 $\pm 5.4\times10^{-6}$	9.9×10^{-4}	9.9×10^{-4}	-	$2.7\times10^{6} \pm 6.1\times10^{3}$	0.029
0.0005	0.1045065 $\pm 2.6\times10^{-6}$	4.8×10^{-4}	4.8×10^{-4}	-	$1.1\times10^{7} \pm 5.0\times10^{3}$	0.028

Table 4: gBM, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.10390 $\pm 1.8\times10^{-4}$	3.4×10^{-2}	3.4×10^{-2}	5.8×10^{-3}	$6.1\times10^{3} \pm 9.4\times10^{0}$	0.078
0.0200	0.104122 $\pm 7.4\times10^{-5}$	1.4×10^{-2}	1.4×10^{-2}	3.7×10^{-3}	$1.5\times10^{4} \pm 5.1\times10^{1}$	0.032
0.0100	0.104175 $\pm 3.8\times10^{-5}$	7.7×10^{-3}	7.0×10^{-3}	3.2×10^{-3}	$4.9\times10^{4} \pm 2.0\times10^{2}$	0.032
0.0050	0.104202 $\pm 1.8\times10^{-5}$	4.4×10^{-3}	3.4×10^{-3}	2.9×10^{-3}	$1.9\times10^{5} \pm 7.8\times10^{2}$	0.041
0.0020	0.1043864 $\pm 7.9\times10^{-6}$	1.8×10^{-3}	1.5×10^{-3}	1.1×10^{-3}	$1.2\times10^{6} \pm 4.1\times10^{3}$	0.046
0.0010	0.1044495 $\pm 4.0\times10^{-6}$	9.1×10^{-4}	7.3×10^{-4}	5.4×10^{-4}	$5.0\times10^{6} \pm 2.1\times10^{4}$	0.046
0.0005	0.1044775 $\pm 2.0\times10^{-6}$	4.5×10^{-4}	3.6×10^{-4}	2.7×10^{-4}	$2.0\times10^{7} \pm 6.7\times10^{4}$	0.045

Table 5: CIR, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.500	0.012038 $\pm 3.0\times10^{-5}$	4.8×10^{-2}	4.8×10^{-2}	-	$3.2\times10^{4} \pm 3.5\times10^{2}$	0.011
0.200	0.012032 $\pm 3.0\times10^{-5}$	4.8×10^{-2}	4.8×10^{-2}	-	$3.2\times10^{4} \pm 2.3\times10^{2}$	0.011
0.100	0.012022 $\pm 3.0\times10^{-5}$	4.8×10^{-2}	4.8×10^{-2}	-	$3.2\times10^{4} \pm 4.4\times10^{2}$	0.011
0.050	0.012050 $\pm 2.8\times10^{-5}$	4.6×10^{-2}	4.5×10^{-2}	-	$3.6\times10^{4} \pm 7.4\times10^{2}$	0.011
0.020	0.012008 $\pm 1.1\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$2.4\times10^{5} \pm 3.4\times10^{3}$	0.011
0.010	0.0120137 $\pm 5.9\times10^{-6}$	9.5×10^{-3}	9.5×10^{-3}	-	$9.0\times10^{5} \pm 5.0\times10^{4}$	0.012
0.005	0.0120121 $\pm 2.6\times10^{-6}$	4.3×10^{-3}	4.2×10^{-3}	-	$4.2\times10^{6} \pm 2.5\times10^{4}$	0.011

Table 6: CIR, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.500	0.01209 $\pm 2.3\times10^{-4}$	3.8×10^{-1}	3.8×10^{-1}	-	$5.3\times10^{3} \pm 9.5\times10^{1}$	0.109
0.200	0.01219 $\pm 1.1\times10^{-4}$	1.8×10^{-1}	1.8×10^{-1}	-	$2.0\times10^{4} \pm 2.9\times10^{2}$	0.095
0.100	0.012023 $\pm 5.4\times10^{-5}$	8.7×10^{-2}	8.7×10^{-2}	-	$9.1\times10^{4} \pm 7.5\times10^{2}$	0.099
0.050	0.012041 $\pm 2.9\times10^{-5}$	4.7×10^{-2}	4.7×10^{-2}	-	$3.1\times10^{5} \pm 2.1\times10^{3}$	0.097
0.020	0.012016 $\pm 1.1\times10^{-5}$	1.7×10^{-2}	1.7×10^{-2}	-	$2.3\times10^{6} \pm 4.3\times10^{3}$	0.095
0.010	0.0120146 $\pm 5.8\times10^{-6}$	9.4×10^{-3}	9.4×10^{-3}	-	$7.9\times10^{6} \pm 1.8\times10^{4}$	0.100
0.005	0.0120122 $\pm 3.1\times10^{-6}$	4.9×10^{-3}	4.9×10^{-3}	-	$2.9\times10^{7} \pm 4.7\times10^{4}$	0.103

Table 7: CIR, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.500	0.01214 $\pm 2.8\times10^{-4}$	4.4×10^{-1}	4.4×10^{-1}	-	$3.5\times10^{3} \pm 5.8\times10^{1}$	0.098
0.200	0.01206 $\pm 1.1\times10^{-4}$	1.8×10^{-1}	1.8×10^{-1}	-	$2.2\times10^{4} \pm 2.9\times10^{2}$	0.101
0.100	0.012026 $\pm 5.6\times10^{-5}$	8.9×10^{-2}	8.9×10^{-2}	-	$8.5\times10^{4} \pm 3.3\times10^{2}$	0.098
0.050	0.012033 $\pm 3.0\times10^{-5}$	4.8×10^{-2}	4.8×10^{-2}	-	$3.1\times10^{5} \pm 3.1\times10^{3}$	0.102
0.020	0.012019 $\pm 1.1\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$2.2\times10^{6} \pm 1.5\times10^{3}$	0.105
0.010	0.0120088 $\pm 6.1\times10^{-6}$	9.7×10^{-3}	9.7×10^{-3}	-	$7.4\times10^{6} \pm 2.0\times10^{4}$	0.101
0.005	0.0120119 $\pm 3.0\times10^{-6}$	4.8×10^{-3}	4.8×10^{-3}	-	$2.9\times10^{7} \pm 3.4\times10^{4}$	0.096

Table 8: CIR, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.500	0.01245 $\pm 2.2\times10^{-4}$	3.5×10^{-1}	3.5×10^{-1}	3.6×10^{-2}	$6.7\times10^{3} \pm 1.0\times10^{1}$	0.118
0.200	0.012351 $\pm 8.3\times10^{-5}$	1.4×10^{-1}	1.3×10^{-1}	2.8×10^{-2}	$2.6\times10^{4} \pm 1.2\times10^{2}$	0.070
0.100	0.012174 $\pm 4.5\times10^{-5}$	7.4×10^{-2}	7.2×10^{-2}	1.3×10^{-2}	$1.1\times10^{5} \pm 8.8\times10^{1}$	0.083
0.050	0.012166 $\pm 2.1\times10^{-5}$	3.7×10^{-2}	3.4×10^{-2}	1.2×10^{-2}	$4.3\times10^{5} \pm 2.9\times10^{2}$	0.082
0.020	0.0120920 $\pm 8.9\times10^{-6}$	1.6×10^{-2}	1.4×10^{-2}	6.3×10^{-3}	$2.8\times10^{6} \pm 1.1\times10^{3}$	0.098
0.010	0.0120447 $\pm 4.5\times10^{-6}$	7.7×10^{-3}	7.3×10^{-3}	2.4×10^{-3}	$1.2\times10^{7} \pm 6.1\times10^{3}$	0.098
0.005	0.0120409 $\pm 2.3\times10^{-6}$	4.3×10^{-3}	3.8×10^{-3}	2.1×10^{-3}	$4.7\times10^{7} \pm 3.1\times10^{4}$	0.124

Table 9: Heston, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.100	0.10441 $\pm$ 4.2×10^{-4}	7.7×10^{-2}	7.7×10^{-2}	-	5.9×10^{3} $\pm$ 2.6×10^{2}	0.384
0.050	0.10435 $\pm$ 2.4×10^{-4}	4.5×10^{-2}	4.5×10^{-2}	-	1.9×10^{4} $\pm$ 5.3×10^{2}	0.414
0.020	0.104573 $\pm$ 9.7×10^{-5}	1.8×10^{-2}	1.8×10^{-2}	-	1.2×10^{5} $\pm$ 2.0×10^{3}	0.433
0.010	0.104612 $\pm$ 4.9×10^{-5}	9.1×10^{-3}	9.1×10^{-3}	-	5.0×10^{5} $\pm$ 9.6×10^{3}	0.455
0.005	0.104592 $\pm$ 2.5×10^{-5}	4.5×10^{-3}	4.5×10^{-3}	-	2.2×10^{6} $\pm$ 2.1×10^{4}	0.495
0.002	0.104596 $\pm$ 1.0×10^{-5}	1.9×10^{-3}	1.9×10^{-3}	-	1.2×10^{7} $\pm$ 9.2×10^{4}	0.454
0.001	0.1045954 $\pm$ 5.3×10^{-6}	9.7×10^{-4}	9.7×10^{-4}	-	4.6×10^{7} $\pm$ 1.6×10^{5}	0.479

Table 10: Heston, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.100	0.10464 $\pm$ 4.3×10^{-4}	7.8×10^{-2}	7.8×10^{-2}	-	8.2×10^{3} $\pm$ 1.3×10^{2}	0.549
0.050	0.10448 $\pm$ 2.6×10^{-4}	4.8×10^{-2}	4.8×10^{-2}	-	2.6×10^{4} $\pm$ 5.0×10^{2}	0.654
0.020	0.10468 $\pm$ 1.0×10^{-4}	1.9×10^{-2}	1.9×10^{-2}	-	1.6×10^{5} $\pm$ 2.4×10^{3}	0.609
0.010	0.104557 $\pm$ 4.7×10^{-5}	8.6×10^{-3}	8.6×10^{-3}	-	7.1×10^{5} $\pm$ 4.6×10^{3}	0.572
0.005	0.104584 $\pm$ 2.6×10^{-5}	4.8×10^{-3}	4.8×10^{-3}	-	2.4×10^{6} $\pm$ 1.6×10^{4}	0.607
0.002	0.1045922 $\pm$ 9.4×10^{-6}	1.7×10^{-3}	1.7×10^{-3}	-	1.8×10^{7} $\pm$ 1.3×10^{5}	0.589
0.001	0.1046015 $\pm$ 5.2×10^{-6}	9.6×10^{-4}	9.6×10^{-4}	-	6.2×10^{7} $\pm$ 2.4×10^{5}	0.621

Table 11: Heston, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.100	0.10493 $\pm$ 4.7×10^{-4}	8.7×10^{-2}	8.7×10^{-2}	-	6.4×10^{3} $\pm$ 1.9×10^{2}	0.531
0.050	0.10459 $\pm$ 2.4×10^{-4}	4.5×10^{-2}	4.5×10^{-2}	-	2.6×10^{4} $\pm$ 3.3×10^{2}	0.568
0.020	0.104636 $\pm$ 9.8×10^{-5}	1.8×10^{-2}	1.8×10^{-2}	-	1.5×10^{5} $\pm$ 1.2×10^{3}	0.542
0.010	0.104589 $\pm$ 4.6×10^{-5}	8.5×10^{-3}	8.5×10^{-3}	-	7.3×10^{5} $\pm$ 6.8×10^{3}	0.584
0.005	0.104591 $\pm$ 2.5×10^{-5}	4.6×10^{-3}	4.6×10^{-3}	-	2.5×10^{6} $\pm$ 9.8×10^{3}	0.595
0.002	0.1045911 $\pm$ 9.2×10^{-6}	1.7×10^{-3}	1.7×10^{-3}	-	1.9×10^{7} $\pm$ 3.2×10^{4}	0.603
0.001	0.1045957 $\pm$ 5.2×10^{-6}	9.6×10^{-4}	9.6×10^{-4}	-	5.8×10^{7} $\pm$ 1.0×10^{5}	0.585

Table 12: Heston, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.100	0.10451 $\pm$ 3.6×10^{-4}	6.7×10^{-2}	6.7×10^{-2}	8.2×10^{-4}	9.9×10^{3} $\pm$ 3.0×10^{1}	0.482
0.050	0.10472 $\pm$ 1.9×10^{-4}	3.5×10^{-2}	3.5×10^{-2}	1.2×10^{-3}	2.7×10^{4} $\pm$ 1.0×10^{2}	0.366
0.020	0.104641 $\pm$ 7.9×10^{-5}	1.5×10^{-2}	1.5×10^{-2}	4.2×10^{-4}	1.7×10^{5} $\pm$ 4.9×10^{2}	0.391
0.010	0.104714 $\pm$ 4.0×10^{-5}	7.5×10^{-3}	7.4×10^{-3}	1.1×10^{-3}	6.8×10^{5} $\pm$ 1.7×10^{3}	0.419
0.005	0.104660 $\pm$ 1.9×10^{-5}	3.5×10^{-3}	3.5×10^{-3}	6.0×10^{-4}	2.7×10^{6} $\pm$ 6.7×10^{3}	0.368
0.002	0.1046499 $\pm$ 7.9×10^{-6}	1.5×10^{-3}	1.5×10^{-3}	5.1×10^{-4}	1.7×10^{7} $\pm$ 5.8×10^{4}	0.451
0.001	0.1046278 $\pm$ 4.1×10^{-6}	8.1×10^{-4}	7.5×10^{-4}	3.0×10^{-4}	7.8×10^{7} $\pm$ 1.4×10^{5}	0.555

Kloeden, P. E., E. Platen, W. Wright. 1992. The approximation of multiple stochastic integrals. *Stochastic Analysis and Applications* **10** 431–441.

Lapeyre, B., E. Temam. 2001. Competitive Monte Carlo methods for the pricing of Asian options. *Journal of Computational Finance* **5**(1) 39–59.

McLeish, D. 2011. A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods and Applications* **17**(4) 301–315.

Rhee, C., P. W. Glynn. 2012. A new approach to unbiased estimation for SDEs. C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, A.M. Uhrmacher, eds., *Proceedings of the 2012 Winter Simulation Conference*.

Wiktorsson, M. 2001. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Annals of Applied Probability* **11**(2) 470–487.

Online Supplement: Additional Computational Results

Here we provide computational results regarding additional path functionals—Asian, lookback, digital, and barrier option—for gBM. We also provide results for the first moment of Vasicek model.

Example 1. *(Asian option for gBM): The SDE for X is*

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

with the parameters selected as $\mu = 0.05$, $\sigma = 0.2$, and $X(0) = 1$. The payoff functional we consider here is an "Asian option" in integral form $f(x) = \exp(-\mu)[\int_0^1 x(t)dt - 1]^+$. We have computed the reference value $\mathbf{E}f(X) = 0.0576310$ using the MLMC estimator with a target RMSE 3.0×10^{-7}. The approximation X_h was simulated with Milstein scheme at the discrete time points, and $f(X_h)$ was computed using Brownian interpolation. See Giles (2008) for detail. Tables 1-4 report the results.

Example 2. *(Lookback option for gBM): We consider the same SDE with the same parameters as in Example 1. Here we consider another path-dependent payoff functional "lookback option" $f(x) = \exp(-\mu)\big(x(1) - \min_{0\leq t\leq 1} x(t)\big)$. The reference value $\mathbf{E}f(X) = 0.172169$ was computed up to RMSE 1.0×10^{-6} using the MLMC estimator. To construct the necessary coupling, we used Milstein scheme with Brownian interpolation; but as in the setting of the antithetic truncated Milstein estimator, we did not construct a sequence of L^2 approximations that converges to $f(X)$. Instead, we have constructed a sequence of differences $\Delta_i' = f(X_{h_i}) - f(X'_{h_{i-1}})$ that converges to 0 in L^2 sufficiently fast with the property that $\mathbf{E}f(X_{h_i}) = \mathbf{E}f(X'_{h_i})$. From this sequence, we have built the unbiased estimators as in Theorem 5. More specifically, X_{h_i} was simulated with Milstein scheme at the discrete time points $(jh_i : j = 0, 1, \cdots, 1/h_i)$ and then $f(X_{h_i})$ was computed using the distribution of the minimum of the Brownian bridge at each discrete time interval. For X'_{h_i}, we first used Milstein scheme to simulate the process at $(jh_i : j = 0, 1, \cdots, 1/h_i)$, and refined the path at $(jh_i + h_i/2 : j = 0, 1, \cdots, 1/h_i)$ using the Brownian bridge interpolation. Given X'_{h_i} at such discrete time points, we have computed $f(X'_{h_i})$ using the distribution of the minimum of the Brownian interpolation between each discrete time points. See Giles (2008) for more detail. Tables 5-8 report the results.*

Example 3. *(Digital option for gBM): Again, we consider the same SDE with the same parameters as in Example 1, and a payoff functional "digital option" $f(x) = \exp(-\mu)\mathbb{I}(x(1) \geq 1)$. The exact value of $\mathbf{E}f(X)$ is 0.532324815. Here the payoff depends only on the final value, but the challenge is that it has a discontinuity, with which the L^2 error associated with the Milstein scheme degrades. To hurdle this difficulty, we have smoothed the payoff using the technique of conditional expectation at the last discrete time step of the Milstein scheme as suggested in Giles (2008). Similarly to* *the Example 2, this approach doesn't produce a sequence of L^2 approximations, but it produces a sequence of differences Δ_i''s with which we can build unbiased estimators applying Theorem 5. Tables 9-12 report the results.*

Example 4. *(Barrier option for gBM): The last payoff functional for gBM we consider is the down-and-out "barrier option" of the form* $f(x) = \exp(-\mu)[x(1)-1]^+ \mathbb{I}(\tau(x) > 1)$ *where* $\tau(x) = \inf_{t\geq 0}\{x(t) < 0.85\}$. *The quantity* $\mathbf{E}f(X) = 0.099491$ *was computed up to RMSE* 1.0×10^{-6} *using the MLMC estimator. Milstein scheme combined with a similar technique as in Example 2 was used to produce a sequence of differences* Δ_i', *and Theorem 5 was used to construct unbiased estimators from this sequence. See Giles (2008) for more detail on the coupling construction. Tables 13-16 report the results.*

Example 5. *(Vasicek model): The SDE for Vasicek model is*

$$dX(t) = \kappa(\theta - X(t))dt + \sigma dB(t),$$

with parameters given by $\mu = 0.05$, $\kappa = 5$, $\theta = 0.04$, $\sigma = 0.05$, *and* $X(0) = 0.04$. *We provide numerical results for the first moment* $f(x) = x$; *the standard Milstein scheme was used to generate* X_h. *The quantity* $\mathbf{E}f(X)$ *is* 0.04. *Tables 17-20 report the results.*

References

Giles, M. B. 2008. Improved multilevel Monte Carlo convergence using the Milstein scheme. A. Keller, S. Heinrich, H. Niederreiter, eds., *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Springer-Verlag, 343–358.

Table 1: gBM, Asian Option, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.05742 $\pm 1.4\times10^{-4}$	4.5×10^{-2}	4.5×10^{-2}	-	$1.5\times10^{3} \pm 2.7\times10^{1}$	0.010
0.0200	0.057661 $\pm 5.3\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$9.9\times10^{3} \pm 9.7\times10^{1}$	0.010
0.0100	0.057607 $\pm 2.8\times10^{-5}$	9.3×10^{-3}	9.3×10^{-3}	-	$3.6\times10^{4} \pm 4.6\times10^{2}$	0.010
0.0050	0.057620 $\pm 1.3\times10^{-5}$	4.3×10^{-3}	4.2×10^{-3}	-	$1.7\times10^{5} \pm 7.8\times10^{2}$	0.010
0.0020	0.0576295 $\pm 5.7\times10^{-6}$	1.9×10^{-3}	1.9×10^{-3}	-	$8.7\times10^{5} \pm 5.8\times10^{3}$	0.010
0.0010	0.0576310 $\pm 2.6\times10^{-6}$	8.8×10^{-4}	8.9×10^{-4}	-	$4.3\times10^{6} \pm 5.2\times10^{3}$	0.011
0.0005	0.0576303 $\pm 1.4\times10^{-6}$	4.8×10^{-4}	4.8×10^{-4}	-	$1.3\times10^{7} \pm 2.5\times10^{4}$	0.010

Table 2: gBM, Asian Option, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.05752 $\pm 1.3\times10^{-4}$	4.4×10^{-2}	4.4×10^{-2}	-	$1.5\times10^{3} \pm 3.1\times10^{1}$	0.010
0.0200	0.057558 $\pm 5.6\times10^{-5}$	1.9×10^{-2}	1.9×10^{-2}	-	$1.0\times10^{4} \pm 9.9\times10^{1}$	0.012
0.0100	0.057620 $\pm 2.8\times10^{-5}$	9.2×10^{-3}	9.2×10^{-3}	-	$3.9\times10^{4} \pm 2.6\times10^{3}$	0.011
0.0050	0.057619 $\pm 1.3\times10^{-5}$	4.4×10^{-3}	4.4×10^{-3}	-	$1.6\times10^{5} \pm 1.8\times10^{3}$	0.010
0.0020	0.0576283 $\pm 5.5\times10^{-6}$	1.8×10^{-3}	1.8×10^{-3}	-	$9.5\times10^{5} \pm 1.8\times10^{3}$	0.011
0.0010	0.0576292 $\pm 2.9\times10^{-6}$	9.6×10^{-4}	9.6×10^{-4}	-	$3.3\times10^{6} \pm 2.3\times10^{4}$	0.010
0.0005	0.0576313 $\pm 1.4\times10^{-6}$	4.8×10^{-4}	4.8×10^{-4}	-	$1.3\times10^{7} \pm 2.2\times10^{4}$	0.010

Table 3: gBM, Asian Option, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.05734 $\pm 1.4\times10^{-4}$	4.6×10^{-2}	4.6×10^{-2}	-	$1.4\times10^{3} \pm 2.1\times10^{1}$	0.010
0.0200	0.057617 $\pm 5.3\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$9.4\times10^{3} \pm 1.0\times10^{2}$	0.010
0.0100	0.057609 $\pm 2.8\times10^{-5}$	9.4×10^{-3}	9.4×10^{-3}	-	$3.3\times10^{4} \pm 2.7\times10^{2}$	0.010
0.0050	0.057629 $\pm 1.4\times10^{-5}$	4.6×10^{-3}	4.6×10^{-3}	-	$1.6\times10^{5} \pm 2.3\times10^{2}$	0.011
0.0020	0.0576257 $\pm 5.7\times10^{-6}$	1.9×10^{-3}	1.9×10^{-3}	-	$8.0\times10^{5} \pm 2.6\times10^{3}$	0.010
0.0010	0.0576307 $\pm 2.6\times10^{-6}$	8.7×10^{-4}	8.7×10^{-4}	-	$4.0\times10^{6} \pm 1.9\times10^{3}$	0.010
0.0005	0.0576327 $\pm 1.5\times10^{-6}$	4.9×10^{-4}	4.9×10^{-4}	-	$1.2\times10^{7} \pm 1.2\times10^{4}$	0.010

Table 4: gBM, Asian Option, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.057388 $\pm 9.7\times10^{-5}$	3.3×10^{-2}	3.2×10^{-2}	4.2×10^{-3}	$6.2\times10^{3} \pm 7.0\times10^{0}$	0.022
0.0200	0.057392 $\pm 4.2\times10^{-5}$	1.5×10^{-2}	1.4×10^{-2}	4.1×10^{-3}	$1.7\times10^{4} \pm 4.2\times10^{1}$	0.012
0.0100	0.057450 $\pm 2.1\times10^{-5}$	7.7×10^{-3}	7.1×10^{-3}	3.1×10^{-3}	$5.7\times10^{4} \pm 1.4\times10^{2}$	0.011
0.0050	0.057472 $\pm 1.0\times10^{-5}$	4.4×10^{-3}	3.4×10^{-3}	2.8×10^{-3}	$2.2\times10^{5} \pm 6.4\times10^{2}$	0.014
0.0020	0.0575651 $\pm 4.3\times10^{-6}$	1.8×10^{-3}	1.4×10^{-3}	1.1×10^{-3}	$1.4\times10^{6} \pm 3.1\times10^{3}$	0.016
0.0010	0.0575994 $\pm 2.3\times10^{-6}$	9.4×10^{-4}	7.7×10^{-4}	5.4×10^{-4}	$5.8\times10^{6} \pm 1.2\times10^{4}$	0.017
0.0005	0.0576153 $\pm 1.1\times10^{-6}$	4.6×10^{-4}	3.7×10^{-4}	2.7×10^{-4}	$2.4\times10^{7} \pm 5.7\times10^{4}$	0.016

Table 5: gBM, Lookback Option, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.17232 $\pm 2.9\times10^{-4}$	3.3×10^{-2}	3.3×10^{-2}	-	$1.6\times10^{3} \pm 4.5\times10^{1}$	0.052
0.0200	0.17218 $\pm 1.7\times10^{-4}$	1.9×10^{-2}	1.9×10^{-2}	-	$5.6\times10^{3} \pm 1.6\times10^{2}$	0.057
0.0100	0.172200 $\pm 8.0\times10^{-5}$	9.0×10^{-3}	9.0×10^{-3}	-	$2.2\times10^{4} \pm 1.1\times10^{3}$	0.053
0.0050	0.172196 $\pm 3.9\times10^{-5}$	4.4×10^{-3}	4.4×10^{-3}	-	$9.2\times10^{4} \pm 6.2\times10^{2}$	0.052
0.0020	0.172177 $\pm 1.7\times10^{-5}$	1.9×10^{-3}	1.9×10^{-3}	-	$5.1\times10^{5} \pm 5.3\times10^{3}$	0.055
0.0010	0.1721723 $\pm 8.1\times10^{-6}$	9.1×10^{-4}	9.1×10^{-4}	-	$2.3\times10^{6} \pm 8.5\times10^{3}$	0.057
0.0005	0.1721645 $\pm 3.8\times10^{-6}$	4.3×10^{-4}	4.3×10^{-4}	-	$8.7\times10^{6} \pm 7.7\times10^{3}$	0.048

Table 6: gBM, Lookback Option, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.17264 $\pm 2.5\times10^{-4}$	2.9×10^{-2}	2.9×10^{-2}	-	$2.0\times10^{3} \pm 2.7\times10^{1}$	0.048
0.0200	0.17222 $\pm 1.7\times10^{-4}$	1.9×10^{-2}	1.9×10^{-2}	-	$5.3\times10^{3} \pm 1.1\times10^{2}$	0.054
0.0100	0.172299 $\pm 7.9\times10^{-5}$	8.9×10^{-3}	8.8×10^{-3}	-	$2.3\times10^{4} \pm 3.8\times10^{2}$	0.054
0.0050	0.172192 $\pm 4.4\times10^{-5}$	4.9×10^{-3}	4.9×10^{-3}	-	$8.1\times10^{4} \pm 8.8\times10^{2}$	0.058
0.0020	0.172150 $\pm 1.6\times10^{-5}$	1.8×10^{-3}	1.8×10^{-3}	-	$5.9\times10^{5} \pm 2.6\times10^{3}$	0.055
0.0010	0.1721705 $\pm 8.1\times10^{-6}$	9.1×10^{-4}	9.1×10^{-4}	-	$2.0\times10^{6} \pm 9.2\times10^{3}$	0.049
0.0005	0.1721755 $\pm 4.0\times10^{-6}$	4.5×10^{-4}	4.5×10^{-4}	-	$8.9\times10^{6} \pm 2.5\times10^{4}$	0.054

Table 7: gBM, Lookback Option, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.17217 $\pm 3.1\times10^{-4}$	3.5×10^{-2}	3.5×10^{-2}	-	$1.5\times10^{3} \pm 5.2\times10^{1}$	0.053
0.0200	0.17228 $\pm 1.7\times10^{-4}$	1.9×10^{-2}	1.9×10^{-2}	-	$5.5\times10^{3} \pm 8.2\times10^{1}$	0.058
0.0100	0.172184 $\pm 8.2\times10^{-5}$	9.1×10^{-3}	9.1×10^{-3}	-	$2.4\times10^{4} \pm 2.6\times10^{2}$	0.060
0.0050	0.172177 $\pm 4.2\times10^{-5}$	4.7×10^{-3}	4.7×10^{-3}	-	$8.6\times10^{4} \pm 7.4\times10^{2}$	0.057
0.0020	0.172160 $\pm 1.5\times10^{-5}$	1.7×10^{-3}	1.7×10^{-3}	-	$6.3\times10^{5} \pm 3.3\times10^{3}$	0.054
0.0010	0.1721715 $\pm 8.2\times10^{-6}$	9.2×10^{-4}	9.2×10^{-4}	-	$2.1\times10^{6} \pm 3.6\times10^{3}$	0.053
0.0005	0.1721673 $\pm 4.4\times10^{-6}$	4.9×10^{-4}	4.9×10^{-4}	-	$8.0\times10^{6} \pm 1.1\times10^{4}$	0.057

Table 8: gBM, Lookback Option, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.17274 $\pm 2.8\times10^{-4}$	3.2×10^{-2}	3.2×10^{-2}	3.3×10^{-3}	$5.2\times10^{3} \pm 2.7\times10^{0}$	0.156
0.0200	0.17282 $\pm 1.2\times10^{-4}$	1.4×10^{-2}	1.3×10^{-2}	3.8×10^{-3}	$9.3\times10^{3} \pm 1.2\times10^{1}$	0.054
0.0100	0.172763 $\pm 6.5\times10^{-5}$	8.0×10^{-3}	7.2×10^{-3}	3.5×10^{-3}	$2.7\times10^{4} \pm 4.1\times10^{1}$	0.052
0.0050	0.172656 $\pm 3.3\times10^{-5}$	4.7×10^{-3}	3.7×10^{-3}	2.8×10^{-3}	$1.1\times10^{5} \pm 2.6\times10^{2}$	0.069
0.0020	0.172330 $\pm 1.3\times10^{-5}$	1.8×10^{-3}	1.5×10^{-3}	9.4×10^{-4}	$7.4\times10^{5} \pm 7.1\times10^{2}$	0.068
0.0010	0.1722467 $\pm 6.2\times10^{-6}$	8.3×10^{-4}	7.0×10^{-4}	4.5×10^{-4}	$3.1\times10^{6} \pm 2.1\times10^{3}$	0.063
0.0005	0.1722030 $\pm 3.2\times10^{-6}$	4.0×10^{-4}	3.5×10^{-4}	2.0×10^{-4}	$1.3\times10^{7} \pm 4.8\times10^{3}$	0.061

Table 9: gBM, Digital Option, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.00500	0.532325 $\pm 6.4\times10^{-5}$	2.3×10^{-3}	2.3×10^{-3}	-	$3.1\times10^{4} \pm 1.1\times10^{3}$	0.048
0.00200	0.532293 $\pm 5.0\times10^{-5}$	1.8×10^{-3}	1.8×10^{-3}	-	$5.0\times10^{4} \pm 2.1\times10^{3}$	0.046
0.00100	0.532329 $\pm 2.5\times10^{-5}$	9.1×10^{-4}	9.1×10^{-4}	-	$2.2\times10^{5} \pm 1.0\times10^{4}$	0.052
0.00050	0.532314 $\pm 1.2\times10^{-5}$	4.5×10^{-4}	4.5×10^{-4}	-	$9.6\times10^{5} \pm 4.0\times10^{4}$	0.056
0.00020	0.5323245 $\pm 5.0\times10^{-6}$	1.8×10^{-4}	1.8×10^{-4}	-	$6.5\times10^{6} \pm 1.5\times10^{5}$	0.060
0.00010	0.5323254 $\pm 2.5\times10^{-6}$	9.1×10^{-5}	9.1×10^{-5}	-	$2.5\times10^{7} \pm 5.1\times10^{5}$	0.059
0.00005	0.5323251 $\pm 1.2\times10^{-6}$	4.2×10^{-5}	4.2×10^{-5}	-	$1.1\times10^{8} \pm 3.8\times10^{6}$	0.057

Table 10: gBM, Digital Option, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.00500	0.532242 $\pm 8.1\times10^{-5}$	2.9×10^{-3}	2.9×10^{-3}	-	$2.9\times10^{4} \pm 2.1\times10^{3}$	0.071
0.00200	0.532247 $\pm 5.1\times10^{-5}$	1.8×10^{-3}	1.8×10^{-3}	-	$7.5\times10^{4} \pm 4.2\times10^{3}$	0.071
0.00100	0.532263 $\pm 2.5\times10^{-5}$	9.3×10^{-4}	9.2×10^{-4}	-	$3.1\times10^{5} \pm 1.2\times10^{4}$	0.076
0.00050	0.532308 $\pm 1.3\times10^{-5}$	4.8×10^{-4}	4.8×10^{-4}	-	$1.3\times10^{6} \pm 3.8\times10^{4}$	0.088
0.00020	0.5323216 $\pm 5.0\times10^{-6}$	1.8×10^{-4}	1.8×10^{-4}	-	$8.9\times10^{6} \pm 2.0\times10^{5}$	0.084
0.00010	0.5323209 $\pm 2.5\times10^{-6}$	9.1×10^{-5}	9.1×10^{-5}	-	$3.9\times10^{7} \pm 6.3\times10^{5}$	0.092
0.00005	0.5323239 $\pm 1.4\times10^{-6}$	4.9×10^{-5}	4.9×10^{-5}	-	$1.4\times10^{8} \pm 3.0\times10^{6}$	0.096

Table 11: gBM, Digital Option, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.00500	0.53197 $\pm 1.1\times10^{-4}$	4.0×10^{-3}	4.0×10^{-3}	-	$1.2\times10^{4} \pm 4.9\times10^{2}$	0.056
0.00200	0.532163 $\pm 5.3\times10^{-5}$	2.0×10^{-3}	1.9×10^{-3}	-	$8.1\times10^{4} \pm 6.3\times10^{3}$	0.088
0.00100	0.532281 $\pm 2.5\times10^{-5}$	8.9×10^{-4}	8.9×10^{-4}	-	$3.4\times10^{5} \pm 1.3\times10^{4}$	0.076
0.00050	0.532309 $\pm 1.3\times10^{-5}$	4.6×10^{-4}	4.6×10^{-4}	-	$1.5\times10^{6} \pm 6.7\times10^{4}$	0.091
0.00020	0.5323213 $\pm 5.2\times10^{-6}$	1.9×10^{-4}	1.9×10^{-4}	-	$9.4\times10^{6} \pm 1.8\times10^{5}$	0.095
0.00010	0.5323224 $\pm 2.4\times10^{-6}$	8.7×10^{-5}	8.6×10^{-5}	-	$3.9\times10^{7} \pm 7.3\times10^{5}$	0.083
0.00005	0.5323241 $\pm 1.3\times10^{-6}$	4.8×10^{-5}	4.8×10^{-5}	-	$1.3\times10^{8} \pm 1.3\times10^{6}$	0.087

Table 12: gBM, Digital Option, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.00500	0.532374 $\pm 9.3\times10^{-5}$	3.4×10^{-3}	3.4×10^{-3}	9.3×10^{-5}	$4.5\times10^{3} \pm 7.3\times10^{1}$	0.015
0.00200	0.532369 $\pm 4.3\times10^{-5}$	1.6×10^{-3}	1.6×10^{-3}	8.2×10^{-5}	$3.1\times10^{4} \pm 3.6\times10^{2}$	0.021
0.00100	0.532401 $\pm 2.2\times10^{-5}$	8.0×10^{-4}	7.9×10^{-4}	1.4×10^{-4}	$1.5\times10^{5} \pm 1.6\times10^{3}$	0.027
0.00050	0.532381 $\pm 1.0\times10^{-5}$	3.9×10^{-4}	3.7×10^{-4}	1.1×10^{-4}	$7.7\times10^{5} \pm 5.4\times10^{3}$	0.032
0.00020	0.5323562 $\pm 4.0\times10^{-6}$	1.6×10^{-4}	1.5×10^{-4}	5.9×10^{-5}	$6.2\times10^{6} \pm 3.8\times10^{4}$	0.043
0.00010	0.5323433 $\pm 2.2\times10^{-6}$	8.6×10^{-5}	7.9×10^{-5}	3.5×10^{-5}	$2.9\times10^{7} \pm 1.8\times10^{5}$	0.062
0.00005	0.5323353 $\pm 1.1\times10^{-6}$	4.3×10^{-5}	3.8×10^{-5}	2.0×10^{-5}	$1.4\times10^{8} \pm 7.8\times10^{5}$	0.071

Table 13: gBM, Barrier Option, Coupled-Sum Unbiased Estimator $\bar{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.09906 $\pm\ 2.4\times10^{-4}$	4.6×10^{-2}	4.6×10^{-2}	-	$2.0\times10^{3} \pm 6.2\times10^{1}$	0.042
0.0200	0.099294 $\pm\ 9.4\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$1.3\times10^{4} \pm 2.9\times10^{2}$	0.045
0.0100	0.099379 $\pm\ 4.7\times10^{-5}$	9.2×10^{-3}	9.2×10^{-3}	-	$5.7\times10^{4} \pm 9.8\times10^{2}$	0.048
0.0050	0.099469 $\pm\ 2.4\times10^{-5}$	4.6×10^{-3}	4.6×10^{-3}	-	$2.2\times10^{5} \pm 4.1\times10^{3}$	0.046
0.0020	0.0994768 $\pm\ 9.0\times10^{-6}$	1.7×10^{-3}	1.7×10^{-3}	-	$1.6\times10^{6} \pm 1.7\times10^{4}$	0.047
0.0010	0.0994848 $\pm\ 4.7\times10^{-6}$	9.1×10^{-4}	9.1×10^{-4}	-	$5.3\times10^{6} \pm 4.7\times10^{4}$	0.043
0.0005	0.0994929 $\pm\ 2.6\times10^{-6}$	5.0×10^{-4}	5.0×10^{-4}	-	$2.1\times10^{7} \pm 1.4\times10^{6}$	0.052

Table 14: gBM, Barrier Option, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.09902 $\pm\ 2.4\times10^{-4}$	4.6×10^{-2}	4.6×10^{-2}	-	$2.1\times10^{3} \pm 1.0\times10^{2}$	0.045
0.0200	0.099366 $\pm\ 9.4\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$1.5\times10^{4} \pm 2.9\times10^{2}$	0.049
0.0100	0.099463 $\pm\ 4.6\times10^{-5}$	9.0×10^{-3}	9.0×10^{-3}	-	$5.4\times10^{4} \pm 8.4\times10^{2}$	0.043
0.0050	0.099467 $\pm\ 2.2\times10^{-5}$	4.3×10^{-3}	4.3×10^{-3}	-	$2.5\times10^{5} \pm 2.9\times10^{3}$	0.047
0.0020	0.0994848 $\pm\ 9.7\times10^{-6}$	1.9×10^{-3}	1.9×10^{-3}	-	$1.4\times10^{6} \pm 2.5\times10^{4}$	0.049
0.0010	0.0994888 $\pm\ 4.4\times10^{-6}$	8.6×10^{-4}	8.6×10^{-4}	-	$6.4\times10^{6} \pm 5.6\times10^{4}$	0.047
0.0005	0.0994910 $\pm\ 2.5\times10^{-6}$	4.8×10^{-4}	4.8×10^{-4}	-	$2.1\times10^{7} \pm 3.1\times10^{5}$	0.049

Table 15: gBM, Barrier Option, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.09881 $\pm\ 2.4\times10^{-4}$	4.6×10^{-2}	4.6×10^{-2}	-	$1.7\times10^{3} \pm 4.0\times10^{1}$	0.036
0.0200	0.099249 $\pm\ 9.2\times10^{-5}$	1.8×10^{-2}	1.8×10^{-2}	-	$1.2\times10^{4} \pm 2.2\times10^{2}$	0.038
0.0100	0.099345 $\pm\ 4.7\times10^{-5}$	9.1×10^{-3}	9.0×10^{-3}	-	$4.8\times10^{4} \pm 1.3\times10^{3}$	0.040
0.0050	0.099478 $\pm\ 2.3\times10^{-5}$	4.5×10^{-3}	4.5×10^{-3}	-	$2.1\times10^{5} \pm 7.8\times10^{3}$	0.042
0.0020	0.0994846 $\pm\ 9.4\times10^{-6}$	1.8×10^{-3}	1.8×10^{-3}	-	$1.4\times10^{6} \pm 1.9\times10^{4}$	0.046
0.0010	0.0994852 $\pm\ 4.8\times10^{-6}$	9.4×10^{-4}	9.4×10^{-4}	-	$5.1\times10^{6} \pm 4.3\times10^{4}$	0.044
0.0005	0.0994913 $\pm\ 2.6\times10^{-6}$	5.0×10^{-4}	5.0×10^{-4}	-	$1.9\times10^{7} \pm 2.0\times10^{5}$	0.046

Table 16: gBM, Barrier Option, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.0500	0.09888 $\pm\ 1.7\times10^{-4}$	3.4×10^{-2}	3.3×10^{-2}	6.1×10^{-3}	$6.4\times10^{3} \pm 9.4\times10^{0}$	0.073
0.0200	0.098854 $\pm\ 7.3\times10^{-5}$	1.6×10^{-2}	1.4×10^{-2}	6.4×10^{-3}	$1.7\times10^{4} \pm 9.4\times10^{1}$	0.041
0.0100	0.098906 $\pm\ 3.8\times10^{-5}$	9.4×10^{-3}	7.3×10^{-3}	5.9×10^{-3}	$6.4\times10^{4} \pm 4.4\times10^{2}$	0.056
0.0050	0.099223 $\pm\ 1.8\times10^{-5}$	4.5×10^{-3}	3.5×10^{-3}	2.7×10^{-3}	$2.8\times10^{5} \pm 1.9\times10^{3}$	0.055
0.0020	0.0993682 $\pm\ 7.3\times10^{-6}$	1.9×10^{-3}	1.4×10^{-3}	1.2×10^{-3}	$1.9\times10^{6} \pm 1.1\times10^{4}$	0.066
0.0010	0.0994333 $\pm\ 3.7\times10^{-6}$	9.2×10^{-4}	7.1×10^{-4}	5.8×10^{-4}	$8.0\times10^{6} \pm 4.5\times10^{4}$	0.067
0.0005	0.0994659 $\pm\ 1.9\times10^{-6}$	4.5×10^{-4}	3.7×10^{-4}	2.5×10^{-4}	$3.4\times10^{7} \pm 1.7\times10^{5}$	0.066

Table 17: Vasicek, First Moment, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.200	0.040017 $\pm$ 3.0×10^{-5}	1.5×10^{-2}	1.5×10^{-2}	-	4.2×10^{4} $\pm$ 5.9×10^{2}	0.014
0.100	0.039979 $\pm$ 3.1×10^{-5}	1.5×10^{-2}	1.5×10^{-2}	-	4.1×10^{4} $\pm$ 4.2×10^{2}	0.015
0.050	0.040031 $\pm$ 3.0×10^{-5}	1.4×10^{-2}	1.4×10^{-2}	-	4.2×10^{4} $\pm$ 8.1×10^{2}	0.014
0.020	0.040012 $\pm$ 3.0×10^{-5}	1.5×10^{-2}	1.5×10^{-2}	-	4.2×10^{4} $\pm$ 4.3×10^{2}	0.014
0.010	0.040000 $\pm$ 2.0×10^{-5}	9.5×10^{-3}	9.5×10^{-3}	-	1.0×10^{5} $\pm$ 1.5×10^{3}	0.015
0.005	0.0400132 $\pm$ 9.0×10^{-6}	4.4×10^{-3}	4.4×10^{-3}	-	4.8×10^{5} $\pm$ 3.2×10^{3}	0.015
0.002	0.0399986 $\pm$ 4.0×10^{-6}	1.9×10^{-3}	1.9×10^{-3}	-	2.5×10^{6} $\pm$ 2.2×10^{4}	0.015

Table 18: Vasicek, First Moment, Independent-Sum Unbiased Estimator $\tilde{Z}$, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.200	0.03968 $\pm$ 3.2×10^{-4}	1.6×10^{-1}	1.6×10^{-1}	-	5.4×10^{3} $\pm$ 4.2×10^{1}	0.212
0.100	0.03992 $\pm$ 1.8×10^{-4}	8.8×10^{-2}	8.8×10^{-2}	-	1.8×10^{4} $\pm$ 2.2×10^{2}	0.228
0.050	0.040034 $\pm$ 9.8×10^{-5}	4.7×10^{-2}	4.7×10^{-2}	-	6.4×10^{4} $\pm$ 6.9×10^{2}	0.229
0.020	0.040001 $\pm$ 3.7×10^{-5}	1.8×10^{-2}	1.8×10^{-2}	-	4.7×10^{5} $\pm$ 4.6×10^{3}	0.237
0.010	0.039979 $\pm$ 2.0×10^{-5}	9.5×10^{-3}	9.5×10^{-3}	-	1.6×10^{6} $\pm$ 5.3×10^{3}	0.228
0.005	0.0399965 $\pm$ 8.8×10^{-6}	4.2×10^{-3}	4.2×10^{-3}	-	8.0×10^{6} $\pm$ 6.3×10^{3}	0.232
0.002	0.0399947 $\pm$ 4.2×10^{-6}	2.0×10^{-3}	2.0×10^{-3}	-	3.6×10^{7} $\pm$ 5.4×10^{4}	0.238

Table 19: Vasicek, First Moment, Single-Term Unbiased Estimator Z, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.200	0.04046 $\pm$ 3.9×10^{-4}	1.9×10^{-1}	1.9×10^{-1}	-	4.3×10^{3} $\pm$ 5.9×10^{1}	0.239
0.100	0.03987 $\pm$ 1.9×10^{-4}	9.0×10^{-2}	9.0×10^{-2}	-	1.7×10^{4} $\pm$ 9.3×10^{1}	0.218
0.050	0.039926 $\pm$ 9.7×10^{-5}	4.7×10^{-2}	4.7×10^{-2}	-	6.4×10^{4} $\pm$ 7.8×10^{2}	0.223
0.020	0.039994 $\pm$ 3.7×10^{-5}	1.8×10^{-2}	1.8×10^{-2}	-	4.3×10^{5} $\pm$ 3.7×10^{3}	0.221
0.010	0.040001 $\pm$ 2.0×10^{-5}	9.7×10^{-3}	9.7×10^{-3}	-	1.5×10^{6} $\pm$ 7.1×10^{3}	0.220
0.005	0.0400062 $\pm$ 9.1×10^{-6}	4.4×10^{-3}	4.4×10^{-3}	-	7.3×10^{6} $\pm$ 3.8×10^{3}	0.226
0.002	0.0400017 $\pm$ 4.0×10^{-6}	1.9×10^{-3}	1.9×10^{-3}	-	3.7×10^{7} $\pm$ 1.7×10^{4}	0.221

Table 20: Vasicek, First Moment, Multilevel Monte Carlo, 1000 Samples

IRE	90% Confidence Interval	RMSE/α	std/α	bias/α	Work	Work×MSE
0.200	0.04007 $\pm$ 2.7×10^{-4}	1.3×10^{-1}	1.3×10^{-1}	1.7×10^{-3}	8.3×10^{3} $\pm$ 9.7×10^{0}	0.230
0.100	0.03993 $\pm$ 1.5×10^{-4}	7.0×10^{-2}	7.0×10^{-2}	1.8×10^{-3}	2.4×10^{4} $\pm$ 2.4×10^{1}	0.189
0.050	0.040053 $\pm$ 7.4×10^{-5}	3.6×10^{-2}	3.6×10^{-2}	1.3×10^{-3}	9.1×10^{4} $\pm$ 5.9×10^{1}	0.184
0.020	0.040000 $\pm$ 3.0×10^{-5}	1.4×10^{-2}	1.4×10^{-2}	1.1×10^{-5}	5.7×10^{5} $\pm$ 2.7×10^{2}	0.185
0.010	0.039997 $\pm$ 1.4×10^{-5}	6.9×10^{-3}	7.0×10^{-3}	8.2×10^{-5}	2.3×10^{6} $\pm$ 9.5×10^{2}	0.176
0.005	0.0399995 $\pm$ 7.3×10^{-6}	3.5×10^{-3}	3.5×10^{-3}	1.2×10^{-5}	9.1×10^{6} $\pm$ 3.9×10^{3}	0.181
0.002	0.0400005 $\pm$ 3.0×10^{-6}	1.4×10^{-3}	1.4×10^{-3}	1.3×10^{-5}	5.7×10^{7} $\pm$ 2.6×10^{4}	0.186