A NEW APPROACH TO UNBIASED ESTIMATION FOR SDE'S

Chang-han Rhee
Peter W. Glynn

Stanford University
Stanford, CA 94305, U.S.A.

ABSTRACT

In this paper, we introduce a new approach to constructing unbiased estimators when computing expectations of path functionals associated with stochastic differential equations (SDEs). Our randomization idea is closely related to multi-level Monte Carlo and provides a simple mechanism for constructing a finite variance unbiased estimator with "square root convergence rate" whenever one has available a scheme that produces strong error of order greater than 1/2 for the path functional under consideration.

1 INTRODUCTION

We have recently developed a general approach to constructing unbiased estimators, given a family of biased estimators. It turns out that the conditions guaranteeing its validity are closely related to those associated with multi-level Monte Carlo methods; see Rhee and Glynn (2012) for details and a more complete discussion of the theory. In this paper, we briefly describe the idea in the setting of computing solutions of stochastic differential equations and provide an initial numerical exploration intended to shed light on the method's potential effectiveness. As we will see below, the conditions under which our estimator produces an algorithm with "square root convergence rate" essentially coincide with the conditions required by multi-level Monte Carlo to converge at the same rate.

In particular, suppose that we wish to compute an expectation of the form $\alpha = Ek(X)$, where $X = (X(t) : t \geq 0)$ is the solution to the SDE

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

$B = (B(t) : t \geq 0)$ is an m-dimensional standard Brownian motion, $\mu : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times m}$, $k : C[0, \infty) \to \mathbb{R}$, and $C[0, \infty)$ is the space of continuous functions mapping $[0, \infty)$ into $\mathbb{R}^d$. The functions μ and σ are "model primitives" specified by the simulationist that are intended to respectively model the state-dependent drift and volatility of the associated process X. In general, the random variable (rv) $k(X)$ cannot be simulated exactly, because the underlying infinite-dimensional object X cannot be generated exactly. Instead, one typically approximates X via a discrete-time approximation $X_h(\cdot)$. For example, the simplest such approximation is the Euler time-stepping algorithm given by

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

that defines X_h at the time points $0, h, 2h, ...$, with X_h defined at intermediate values via (for example) linear interpolation. Because (2) is only an approximation to the dynamics represented by (1), the rv $k(X_h)$ is only an approximation to $k(X)$, and consequently $k(X_h)$ is a biased estimator for the purpose of computing α. The traditional means of dealing with this is to intelligently select the step size h and number of independent replications R as a function of the computational budget c, so as to maximize the rate of convergence. However, as pointed out by Duffie and Glynn (1995), such biased numerical schemes inevitably lead to Monte Carlo estimators for α that exhibit slower convergence rates than the "canonical" order $c^{-1/2}$ rate associated with Monte Carlo in the presence of unbiased finite variance estimators.

However, Giles (2008a) introduced an intriguing multi-level idea to deal with such biased settings that can dramatically improve the rate of convergence and can even, in some settings, achieve the canonical "square root" convergence rate associated with unbiased Monte Carlo. His approach does not construct an unbiased estimator, however. Rather, the idea is to construct a family of estimators (indexed by the desired error tolerance ε) that has controlled bias. In this paper, we show how it is possible, in a similar computational setting, to go one step further and to produce (exactly) unbiased estimators. The remainder of this paper is organized as follows: We discuss the idea in Section 2 of this paper, while Section 3 is devoted to an initial computational exploration of this approach.

2   THE BASIC IDEA

We consider here a sequence $(X_{h_n} : n \geq 0)$ of discrete-time time-stepping approximations to X that are all constructed on a common probability space in such a way that:

  i) $Ek(X_{h_n}) = Ek(X) + O(h_n)$ as $h_n \to 0$;
  ii) $E|k(X_{h_n}) - k(X)|^2 = O(h_n^{2r})$ as $h_n \to 0$

for some $r > 0$, where $O(f(n))$ represents a function which is bounded by some constant multiple of $f(\cdot)$ as $h_n \to 0$. Assuming, as is often the case for such discretization schemes, that the scheme generates normal rv's that are intended to mirror the Brownian increments of the process B driving the SDE (as in the Euler scheme (2) above), the easiest way to algorithmically obtain an approximating sequence X_{h_n} to X in which the X_{h_n}'s are jointly defined on the same probability space is by successive binary refinement, so that $h_n = 2^{-n}$. In this setting, the new Brownian motion values $(B(j2^{-(n+1)}): j \text{ odd})$ required at discretization $2^{-(n+1)}$ can be obtained from the existing values $(B(j2^{-n}) : j \geq 0)$ by generating $B((2i+1)2^{-(n+1)})$ from its conditional distribution given $B(i2^{-n})$ and $B((i+1)2^{-n})$. On the other hand, one's ability to obtain i) and ii) depends both on the path functional k and on one's choice of discretization scheme.

In particular, if k is a Lipschitz functional on $C[0, \infty)$ (e.g., $k(x) = g(x(1))$ for some Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$), condition ii) is implied by

$$E \sup\{|X_h(jh) - X(jh)|^2 : 0 \leq j \leq \lfloor t/h \rfloor\} = O(h^{2r}),$$

which in turn is closely related to establishing that the discretization X_h exhibits strong order r; see Kloeden and Platen (1992), p.323, for the definition. In addition, if k is further assumed to be smooth with $|k(X)|$ integrable, then i) is satisfied whenever the discretization X_h is known to be of weak order 1 or higher; see Kloeden and Platen (1992), p.327, for the definition. It should be noted that these conditions are (very) closely related to those that appear in the literature on multi-level Monte Carlo for SDEs.

Note that each of the $k(X_{2^{-n}})$'s is a biased estimator for $\alpha = Ek(X)$. To obtain an unbiased estimator, observe that ii) implies the existence of $p > 0$ such that

$$\sum_{n=1}^{\infty} E 2^{np} |k(X_{2^{-n}}) - k(X_{2^{-(n-1)}})|^2 < \infty.$$

Consequently,

$$\sum_{n=1}^{\infty} 2^{np} |k(X_{2^{-n}}) - k(X_{2^{-(n-1)}})|^2 < \infty \quad a.s.$$

from which it follows that

$$|k(X_{2^{-n}}) - k(X_{2^{-(n-1)}})|^2 = O(2^{-np}) \quad a.s.$$

as $n \to \infty$, and hence (in view of ii),

$$k(X) = k(X_1) + \sum_{n=1}^{\infty} \left(k(X_{2^{-n}}) - k(X_{2^{-(n-1)}}) \right). \tag{3}$$

(It has been pointed out to us that this representation for the rv $k(X)$ as an infinite "collapsing sum" was previously and independently developed by McLeish (2011) in a paper that also focuses on construction of unbiased estimators, but which does not apply this idea to SDEs in the way suggested here.)

Because the infinite sum (3) requires infinite work to compute, we use a randomization idea to create a "finite work" estimator with the same expectation. Prior applications of this idea can be found in Glynn (1983) and Fox and Glynn (1989). Specifically, we now introduce a rv N, independent of B, that takes values in the positive integers and has a distribution with unbounded support (so that $P(N \geq n) > 0$ for $n \geq 1$). For such a rv N,

$$Ek(X) = Ek(X_1) + \sum_{n=1}^{\infty} E\left[(k(X_{2^{-n}}) - k(X_{2^{-(n-1)}}))I(N \geq n) \right] / P(N \geq n)$$

$$= E\left[k(X_1) + \sum_{n=1}^{N} (k(X_{2^{-n}}) - k(X_{2^{-(n-1)}}))/P(N \geq n) \right]$$

$$\triangleq EZ.$$

Note that Z is an unbiased estimator for α. This suggests computing α by generating iid replicates of the rv Z. Of course, the "square root" convergence rate of such an estimator is not guaranteed. Given the role that finiteness of the variance plays in obtaining such convergence rates, we next study this issue.

Set $\Delta_i = k(X_{2^{-i}}) - k(X_{2^{-(i-1)}})$ for $i \geq 1$ and observe that

$$EZ^2 = Ek^2(X_1) + E\sum_{i=1}^{N} \Delta_i^2/P(N \geq i)^2 + 2Ek(X_1)\sum_{i=1}^{N} \Delta_i/P(N \geq i) + 2E\sum_{i=1}^{N}\sum_{j=i+1}^{N} \Delta_i\Delta_j/(P(N \geq i)P(N \geq j))$$

$$= Ek^2(X_1) + \sum_{i=1}^{\infty} E\Delta_i^2/P(N \geq i) + 2Ek(X_1)\sum_{i=1}^{\infty} \Delta_i + 2\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E\Delta_i\Delta_j/P(N \geq i)$$

$$= Ek^2(X_1) + \sum_{i=1}^{\infty} E\Delta_i^2/P(N \geq i) + 2Ek(X_1)(k(X) - k(X_1)) + 2\sum_{i=1}^{\infty} E\Delta_i(k(X) - k(X_{2^{-i}}))/P(N \geq i)$$

$$\leq Ek^2(X_1) + 2Ek(X_1)(k(X) - k(X_1)) + \sum_{i=1}^{\infty} O(2^{-2ri})/P(N \geq i),$$

so that $\mathrm{var}Z < \infty$ if $P(N \geq i) \sim c2^{-\gamma i}$ as $i \to \infty$, for $0 < \gamma < 2r$ (where $a_i \sim b_i$ means that $a_i/b_i \to 1$ as $i \to \infty$).

Finally, Glynn and Whitt (1992) prove that "square root convergence rate" ensues if $\mathrm{var}Z < \infty$ and if the expected computational effort required per replication of Z is finite. The expected computational "work" required for each Z is (roughly) given by

$$E\sum_{i=0}^{N} t_i,$$

where t_i is the incremental effort required to compute $k(X_{2^{-i}})$ (given $k(X_1), \ldots, k(X_{2^{-(i-1)}})$), and hence can be expressed as

$$\sum_{i=0}^{\infty} t_i P(N \geq i). \tag{4}$$

An approximation to t_i is $t_i = 2^{i-1}$ (the number of additional Gaussian rv's needed to generate $X_{2^{-i}}$). In order that (4) be finite, we require that $\gamma > 1$. Consequently, a square root convergence rate is ensured when $2r > 1$ (in which case we can, for example, choose $\gamma = (1 + 2r)/2$).

3 A PRELIMINARY COMPUTATIONAL INVESTIGATION

In this section, we implement our method and compare it to the multilevel Monte Carlo algorithm suggested in Giles (2008a). We consider two examples:

Geometric Brownian Motion (GBM): The process under consideration here is the solution to

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

subject to $X(0) = 1$, $\mu = 0.05$, $\sigma = 0.2$, the functional k is $k(x) = x(1)$. For this set of parameters,

$$Ek(X) = 1.05127.$$

Cox-Ingersoll-Ross process (CIR): Here, X solves

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

subject to $X(0) = 0.04$, $\kappa = 5$, $\theta = 0.04$, and $\sigma = 0.25$. The functional k is taken here to be $k(x) = x(1)$. For this example,

$$Ek(X) = 0.04.$$

The numerical scheme used to solve each of the above SDE's was the Milstein scheme; see Kloeden and Platen (1992). For the above problems, we expect $r = 1$. For the purpose of this paper, we do not try to optimize the distribution of N, and instead choose N so that

$$P(N \geq i) = 2^{-3(i-1)/2}$$

for $i \geq 1$. (In other words, we choose γ as the midpoint between 1 and $2r$, although any choice in $(1, 2)$ would provide "square root convergence rate".)

To compare our method to the multi-level Monte Carlo (MLMC) method, we take the view (as in Giles 2008a) that the root mean square error (RMSE) ε to be achieved by the algorithm is given. Giles (2008a) provides a complete description of how to construct a MLMC estimator achieving approximate RMSE ε, and Giles (2008b) demonstrates numerically that the square root convergence rate can be achieved by MLMC with the Milstein scheme. We have used the MATLAB code provided on the author's webpage for Giles (2008b). For our unbiased estimator, we generate independent and identically distributed (iid) replications of the rv Z until such time as the approximate RMSE is less than or equal to ε. In other words, our estimator for α is

$$\frac{1}{N(\varepsilon)}\sum_{i=1}^{N(\varepsilon)} Z_i, \tag{5}$$

where the Z_i's are iid replicates of Z, and

$$N(\varepsilon) = \inf\left\{ n > 10 : \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(Z_i - \frac{1}{n}\sum_{j=1}^{n} Z_j \right)^2 \leq \varepsilon^2 \right\}$$

is the first time that the sample RMSE of the sample mean drops below ε. We use the stopping rule $N(\varepsilon)$ in order to permit easy computation for our estimator, although its use is somewhat unnatural for our estimator (since its use induces bias in our estimator).

Table 1: Unbiased estimation for GBM.

IRE	90% Confidence Interval	RMSE	Work
25%	$1.049 \pm 1.1 \times 10^{-2}$	6.8×10^{-2}	$4.98 \times 10^1 \pm 9.0 \times 10^0$
10%	$1.049 \pm 1.1 \times 10^{-2}$	6.7×10^{-2}	$4.99 \times 10^1 \pm 9.1 \times 10^0$
5%	$1.0454 \pm 9.5 \times 10^{-3}$	5.8×10^{-2}	$7.32 \times 10^1 \pm 1.1 \times 10^1$
2%	$1.0522 \pm 3.1 \times 10^{-3}$	1.9×10^{-2}	$4.40 \times 10^2 \pm 2.9 \times 10^1$
1%	$1.0508 \pm 1.6 \times 10^{-3}$	9.8×10^{-3}	$1.725 \times 10^3 \pm 5.9 \times 10^1$
0.5%	$1.05097 \pm 8.3 \times 10^{-4}$	5.1×10^{-3}	$7.38 \times 10^3 \pm 4.2 \times 10^2$
0.2%	$1.05103 \pm 3.2 \times 10^{-4}$	1.9×10^{-3}	$4.498 \times 10^4 \pm 6.6 \times 10^2$

Table 2: MLMC estimation for GBM.

IRE	90% Confidence Interval	RMSE	Work
25%	$1.0486 \pm 3.4 \times 10^{-3}$	2.1×10^{-2}	$3.100 \times 10^3 \pm 0.0 \times 10^0$
10%	$1.0488 \pm 3.4 \times 10^{-3}$	2.1×10^{-2}	$3.100 \times 10^3 \pm 0.0 \times 10^0$
5%	$1.0517 \pm 3.4 \times 10^{-3}$	2.1×10^{-2}	$3.100 \times 10^3 \pm 0.0 \times 10^0$
2%	$1.0544 \pm 2.4 \times 10^{-3}$	1.5×10^{-2}	$3.204 \times 10^3 \pm 3.6 \times 10^0$
1%	$1.0523 \pm 1.2 \times 10^{-3}$	7.7×10^{-3}	$3.825 \times 10^3 \pm 1.0 \times 10^1$
0.5%	$1.05090 \pm 6.2 \times 10^{-4}$	3.8×10^{-3}	$6.313 \times 10^3 \pm 2.7 \times 10^1$
0.2%	$1.05106 \pm 2.4 \times 10^{-4}$	1.5×10^{-3}	$2.470 \times 10^4 \pm 1.8 \times 10^2$

For each of our two examples, we provide two tables. The first table for each example concerns our new estimator (5); IRE stands for "intended relative error", and $q\%$ then corresponds to setting $\varepsilon = (q/100)|\alpha|$. The 90% confidence interval is then obtained by taking 100 replications of (5) for a given ε, computing the sample mean and sample standard deviation of the 100 observations, and constructing a confidence interval based on the normal approximation. The column corresponding to RMSE is the square root of the average, over the 100 observations, of the square of (5) minus EZ. (Thus, RMSE is reporting the actual root mean square error of the estimator, rather than the intended RMSE that the estimator has been designed to attain asymptotically.) The final column, denoted "work", reports a 90% confidence interval for the expected number of normal rv's generated to construct (5), based on our 100 samples. The second table for each example provides a corresponding set of values for the MLMC estimator.

Taking the amount of work required for computation into account, our results are reasonably comparable to those associated with MLMC, despite the fact that we have done essentially no tuning to optimize the distribution of N for the purposes of this paper. In particular, for the CIR example, our estimator appears to require less work for comparable accuracy, while for the GBM example, MLMC beats our estimator. In addition, our estimator is (arguably) easier to implement than MLMC, since (in its current form) there are no algorithmic parameters that are estimated "on the fly" within the algorithm (in contrast to MLMC). Thus, the unbiased estimators introduced here, despite the independent randomization associated with N (that inflates variance relative to the non-randomized "infinite work" alternative involving the infinite sum (3)), offer a promising computational alternative to MLMC in the presence of SDE numerical schemes having a strong order greater than 1/2.

ACKNOWLEDGMENTS

Table 3: Unbiased estimation for CIR.

IRE	90% Confidence Interval	RMSE	Work
25%	$0.0406 \pm 1.6 \times 10^{-3}$	9.8×10^{-3}	$2.00 \times 10^{2} \pm 2.7 \times 10^{1}$
10%	$0.04045 \pm 6.5 \times 10^{-4}$	3.9×10^{-3}	$1.419 \times 10^{3} \pm 7.1 \times 10^{1}$
5%	$0.04001 \pm 3.4 \times 10^{-4}$	2.0×10^{-3}	$6.42 \times 10^{3} \pm 2.9 \times 10^{2}$
2%	$0.04007 \pm 1.3 \times 10^{-4}$	8.1×10^{-4}	$4.168 \times 10^{4} \pm 1.3 \times 10^{3}$
1%	$0.039938 \pm 6.4 \times 10^{-5}$	4.0×10^{-4}	$1.659 \times 10^{5} \pm 2.8 \times 10^{3}$
0.5%	$0.039988 \pm 3.4 \times 10^{-5}$	2.0×10^{-4}	$6.613 \times 10^{5} \pm 6.3 \times 10^{3}$
0.2%	$0.040014 \pm 1.4 \times 10^{-5}$	8.4×10^{-5}	$4.154 \times 10^{6} \pm 3.9 \times 10^{4}$

Table 4: MLMC estimation for CIR.

IRE	90% Confidence Interval	RMSE	Work
25%	$0.04054 \pm 9.7 \times 10^{-4}$	5.9×10^{-3}	$3.731 \times 10^{3} \pm 1.3 \times 10^{1}$
10%	$0.04019 \pm 4.1 \times 10^{-4}$	2.5×10^{-3}	$1.1519 \times 10^{4} \pm 4.5 \times 10^{1}$
5%	$0.03988 \pm 2.1 \times 10^{-4}$	1.3×10^{-3}	$4.250 \times 10^{4} \pm 1.4 \times 10^{2}$
2%	$0.03986 \pm 1.0 \times 10^{-4}$	6.2×10^{-4}	$2.6595 \times 10^{5} \pm 5.4 \times 10^{2}$
1%	$0.040005 \pm 4.4 \times 10^{-5}$	2.7×10^{-4}	$1.0987 \times 10^{6} \pm 6.9 \times 10^{3}$
0.5%	$0.040008 \pm 2.3 \times 10^{-5}$	1.4×10^{-4}	$4.5729 \times 10^{6} \pm 5.2 \times 10^{3}$
0.2%	$0.039997 \pm 1.0 \times 10^{-5}$	5.8×10^{-5}	$2.8609 \times 10^{7} \pm 2.8 \times 10^{4}$

REFERENCES

Duffie, D., and P. W. Glynn. 1995. "Efficient Monte Carlo Simulation of Security Prices". *Annals of Applied Probability* 5 (4): 897–905.

Fox, B. L., and P. W. Glynn. 1989. "Simulating Discounted Costs". *Management Science* 35:1297–1325.

Giles, M. B. 2008a. "Multilevel Monte Carlo Path Simulation". *Operations Research* 56 (3): 607–617.

Giles, M. B. 2008b. "Improved Multilevel Monte Carlo Convergence Using the Milstein Scheme". In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, edited by A. Keller, S. Heinrich, and H. Niederreiter, 343–358: Springer-Verlag.

Glynn, P. W. 1983. "Randomized Estimators for Time Integrals". Technical report, Mathematics Research Center, University of Wisconsin, Madison.

Glynn, P. W., and W. Whitt. 1992. "The Asymptotic Efficiency of Simulation Estimators". *Operations Research* 40:505–520.

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

McLeish, D. 2011. "A General Method for Debiasing a Monte Carlo Estimator". *Monte Carlo Methods and Applications* 17 (4): 301–315.

Rhee, C., and P. W. Glynn. 2012. "Improved Convergence Rates for Biased Samplers". Working Paper.

AUTHOR BIOGRAPHIES

CHANG-HAN RHEE is currently a Ph.D. student in the Institute for Computational and Mathematical Engineering at Stanford University. He graduated with B.Sc. in the Departments of Mathematics and Computer Science at Seoul National University, South Korea. His research interests include simulation, computational probability, sensitivity analysis and stochastic control. His email address is chrhee@stanford.edu and his web page is http://www.stanford.edu/~chrhee/.

PETER W. GLYNN is currently the Chair of the Department of Management Science and Engineering at Stanford University. He received his Ph.D in Operations Research from Stanford University in 1982. He

then joined the faculty of the University of Wisconsin at Madison, where he held a joint appointment between the Industrial Engineering Department and Mathematics Research Center, and courtesy appointments in Computer Science and Mathematics. In 1987, he returned to Stanford, where he joined the Department of Operations Research. He is now the Thomas Ford Professor of Engineering in the Department of Management Science and Engineering, and also holds a courtesy appointment in the Department of Electrical Engineering. From 1999 to 2005, he served as Deputy Chair of the Department of Management Science and Engineering, and was Director of Stanford's Institute for Computational and Mathematical Engineering from 2006 until 2010. He is a Fellow of INFORMS and a Fellow of the Institute of Mathematical Statistics, has been co-winner of Best Publication Awards from the INFORMS Simulation Society in 1993 and 2008, was a co-winner of the Best (Biannual) Publication Award from the INFORMS Applied Probability Society in 2009, and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. In 2012, he was elected to the National Academy of Engineering. His research interests lie in simulation, computational probability, queueing theory, statistical inference for stochastic processes, and stochastic modeling. His email address is glynn@stanford.edu and his web page is http://www.stanford.edu/~glynn/.