Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

LOW BIAS INTEGRATED PATH ESTIMATORS

James M. Calvin

Department of Computer Science
New Jersey Institute of Technology
Newark, New Jersey, 07102, U.S.A.

ABSTRACT

We consider the problem of estimating the time-average variance constant for a stationary process. A previous paper described an approach based on multiple integrations of the simulation output path, and described the efficiency improvement that can result compared with the method of batch means (which is a special case of the method). In this paper we describe versions of the method that have low bias for moderate simulation run lengths. The method constructs an estimator based on applying a quadratic function to the simulation output. The particular quadratic form is chosen to minimize variance subject to constraints on the order of the bias. Estimators that are first-order and second-order unbiased are described.

1 INTRODUCTION

We consider the steady-state simulation output analysis problem for a stationary process. Suppose a simulation produces output $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)$. Our goal is to estimate

$$\sigma^2 = \lim_{n\to\infty} n \; Var\left(\sum_{i=1}^{n} Y_i/n\right).$$

One approach is to apply a quadratic function to the output vector $\mathbf{Y}_n$. In general, the cost of computing such an estimator is proportional to the square of the simulation run length n. An example of this approach is the batch means estimator, though in the case of batch means the complexity is linear in run length n due to the special form of the quadratic function.

Our approach is as follows. Choose a nonnegative parameter k that is small compared with the square root of the simulation run length n. As the simulation runs, maintain a vector $\widetilde{\mathbf{W}}_n \in \mathbb{R}^{k+1}$ after n simulated steps. At any time n we can construct our variance estimator V_n from $\widetilde{\mathbf{W}}_n$ in time $\mathscr{O}(k^2)$ by applying a quadratic function to $\widetilde{\mathbf{W}}_n$. This way the complexity is reduced to $\mathscr{O}(k^2) = \mathscr{O}(n)$ per computation of the variance estimator.

A previous paper (Calvin 2007), introduced a class of estimators that are based on multiple integrations of the simulated path; the rth component of the vector $\widetilde{\mathbf{W}}_n$ is the r-fold integrated path of the simulation output. In that paper, the quadratic function was chosen so that the limiting distribution of the estimators could be worked out easily and we were mainly concerned with the *efficiency* of the estimators, where the efficiency of an estimator is defined as the reciprocal of the product of mean-squared error and computation time. Numerical experiments showed that with the particular choice of quadratic function used there, the efficiency of the resulting estimator could be an order of magnitude higher than the efficiency of the batch means estimator. However, in experiments with short run lengths it was found that the bias of the estimators was quite large compared with the method of batch means. That finding motivated the present paper, which focuses on bias properties.

2 SETUP AND ASSUMPTIONS

Suppose that a simulation generates a real-valued sequence $Y_1, Y_2, \ldots$. We assume that (Y_i) is a stationary process with $EY_i = \mu$ and that the series

$$\sigma^2 = E(Y_1-\mu)^2 + \sum_{n>1} E\left((Y_1-\mu)(Y_n-\mu)\right)$$

converges absolutely with $\sigma^2 > 0$. For the bias approximations, we will assume the absolute convergence of the sums

$$\lambda_j = \sum_{n>1} n^j E\left((Y_1-\mu)(Y_n-\mu)\right)$$

for $j \le 2$.

1-4244-1306-0/07/$25.00 ©2007 IEEE

3 SIMULATION ALGORITHM

We now outline the proposed method for estimating the parameter σ^2 of the simulated process. Choose a parameter $k \geq 0$ (the integration count parameter). Run a simulation, producing output $\{Y_1, Y_2, \ldots\}$. Define $\widetilde{W}_0^j = 0$ for $0 \leq j \leq k$ and for $i > 0$ set

$$\widetilde{W}_i^0 = \sum_{l=1}^{i} Y_l$$

and for $1 \leq j \leq k$ set

$$\widetilde{W}_i^j = j \sum_{l=1}^{i} \widetilde{W}_l^{j-1}.$$

The data maintained by the simulation method after n simulated steps is $\widetilde{\mathbf{W}}_n = (\widetilde{W}_n^0, \widetilde{W}_n^1, \ldots, \widetilde{W}_n^k)$. (We will use bold typeface for vectors and matrices.) The vector $\widetilde{\mathbf{W}}_n$ can be updated in time $\mathscr{O}(k)$ at each simulation step.

Define

$$W_n^0(t) = n^{-1/2} \sum_{i=1}^{\lceil nt \rceil} Y_i, \quad 0 \leq t \leq 1,$$

and for $r \geq 1$ and $0 \leq t \leq 1$,

$$W_n^r(t) = \int_{s=0}^{t} W_n^{r-1}(s) ds = r \int_{s=0}^{t} W_n^0(s)(t-s)^{r-1} ds.$$

Set $\mathbf{W}_n(t) = \left(W_n^0(t), \ldots, W_n^k(t)\right)$ and $\widetilde{\mathbf{W}}_n(t) = \left(\widetilde{W}_n^0(t), \ldots, \widetilde{W}_n^k(t)\right)$, $0 \leq t \leq 1$.

Define the matrices $\mathbf{A}$ and $\mathbf{N}_n$ by

$$A(r,q) = \begin{cases} (-1)^{r-q} \left[r \atop q \right], & q \leq r, \\ 0 & q > r, \end{cases}$$

where the $\left[r \atop j \right]$ are the Stirling numbers of the first kind (Knuth 1997), and

$$N_n(r,q) = \begin{cases} n^{q+1/2}, & q = r, \\ 0 & q \neq r. \end{cases}$$

The following theorem provides the means to transform the discrete iterated sums into the iterated integrals.

Theorem 1.

$$\mathbf{W}_n(1) = (\mathbf{A}\mathbf{N}_n)^{-1} \widetilde{\mathbf{W}}_n(1).$$

In order to center the output, define

$$\overline{W}_n^0(s) = W_n^0(s) - sW_n^0(1), \quad 0 \leq s \leq 1,$$

and for $r \geq 1$ and $0 \leq t \leq 1$,

$$\overline{W}_n^r(t) = r \int_{s=0}^{t} \overline{W}_n^0(s)(t-s)^{r-1} ds.$$

Set $\overline{\mathbf{W}}_n(t) = \left(\overline{W}_n^0(t), \ldots, \overline{W}_n^k(t)\right)$, $0 \leq t \leq 1$. Define

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1/2 & 0 & \cdots & 0 \\ 1/3 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 1/(k+1) & 0 & \cdots & 0 \end{pmatrix}.$$

Then

$$\overline{\mathbf{W}}_n(1) = (\mathbf{I} - \mathbf{R}) (\mathbf{A}\mathbf{N}_n)^{-1} \widetilde{\mathbf{W}}_n(1).$$

Our output analysis method is based on the following limit theorem.

Theorem 2. *As $n \to \infty$,*

$$\left(\overline{W}_n^1(1), \ldots, \overline{W}_n^k(1)\right) \xrightarrow{\mathscr{D}} \sigma \mathscr{N}(0, \mathbf{C}), \tag{1}$$

where $\mathbf{C} = (C_{ij})$ is the matrix defined by

$$C_{ij} = \frac{ij}{(i+1)(j+1)(i+j+1)}, \quad 1 \leq i,j \leq k.$$

The proof of Theorem 2 follows from Theorem 1 in Calvin (2007), with a different covariance matrix due to a slightly different definition of the W_n^r.

4 BIAS AND VARIANCE

To prepare for the construction of low bias estimators, we will need an approximation to the expectation of products of the $\{\overline{W}_n^r\}$. The following theorem approximates the expectation to order n^{-2}.

Theorem 3. *For $1 \leq \alpha, \beta \leq k$,*

$$E\left(\overline{W}_n^\alpha \overline{W}_n^\beta\right) = \frac{\alpha\beta\sigma^2}{(\alpha+1)(\beta+1)(\alpha+\beta+1)} + \frac{D_1^{\alpha,\beta}}{n} + \frac{D_2^{\alpha,\beta}}{n^2} + o\left(n^{-2}\right) \tag{2}$$

as $n \to \infty$, where

$$D_1^{\alpha,\beta} = \frac{(\alpha\beta - 1)\sigma^2}{2(\alpha+1)(\beta+1)} - \frac{(\alpha\beta+1)\lambda_1}{(\alpha+1)(\beta+1)}, \qquad (3)$$

and

$$\begin{aligned} D_2^{\alpha,\beta} &= \frac{\sigma^2}{12}\left(\alpha+\beta - \frac{\alpha}{\beta+1}I_{(\alpha>1)} - \frac{\beta}{\alpha+1}I_{(\beta>1)}\right) \\ &+ \lambda_1\left(\frac{I_{(\alpha=1)} + \frac{1}{2}\alpha I_{(\alpha>1)}}{\beta+1} + \frac{I_{(\beta=1)} + \frac{1}{2}\beta I_{(\beta>1)}}{\alpha+1} - \frac{\alpha+\beta}{2}\right) \\ &+ \frac{\lambda_2}{2}\left(\frac{\alpha(\alpha-1)+\beta(\beta-1)}{(\alpha+\beta-1)} - \frac{\alpha I_{(\alpha>1)}}{\beta+1} - \frac{\beta I_{(\beta>1)}}{\alpha+1}\right). \end{aligned}$$

We now consider estimators of the form $V_n = \overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}_n$ for a symmetric positive semidefinite matrix $\mathbf{B}$ of order k. We will consider different matrices $\mathbf{B}$, chosen to optimize different objectives. In general, we will choose $\mathbf{B}$ to minimize the asymptotic variance of V_n, subject to constraints on the bias. We next give the bias expansion in terms of the matrix $\mathbf{B}$.

Theorem 4. *Suppose that the matrix* $\mathbf{B}$ *is chosen so that*

$$\sum_{i=0}^{k}\sum_{j=0}^{k} \frac{ij}{(i+1)(j+1)(i+j+1)} B_{ij} = 1. \qquad (4)$$

If in addition we choose $\mathbf{B}$ *such that*

$$\sum_{i=1}^{k}\sum_{j=1}^{k} \frac{ij+1}{(i+1)(j+1)} B_{ij} = 0, \qquad (5)$$

then

$$E\overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}_n = \sigma^2\left(1 + \frac{\beta_1}{n}\right) + o\left(n^{-1}\right),$$

where

$$\beta_1 = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \frac{ij-1}{(i+1)(j+1)} B_{ij}.$$

If, in addition to (4) *and* (5)*, the following constraints hold:*

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\frac{I_{(i=1)} + \frac{1}{2}iI_{(i>1)}}{j+1} + \frac{I_{(j=1)} + \frac{1}{2}jI_{(j>1)}}{i+1} - \frac{i+j}{2}\right) B_{ij} = 0 \qquad (6)$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\frac{i(i-1)+j(j-1)}{(i+j-1)} - \frac{iI_{(i>1)}}{j+1} - \frac{jI_{(j>1)}}{i+1}\right) B_{ij} = 0, \qquad (7)$$

then

$$E\overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}_n = \sigma^2\left(1 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2}\right) + o\left(n^{-2}\right),$$

where

$$\beta_2 = \frac{1}{12}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(i+j - \frac{i}{j+1}I_{(i>1)} - \frac{j}{i+1}I_{(j>1)}\right) B_{ij}.$$

Suppose that the matrix $\mathbf{B}$ satisfies (4) and (5), and define

$$V_n^1 = \frac{\overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}}{1 + \beta_1/n}. \qquad (8)$$

Then from Theorem 4,

$$E\left(V_n^1 - \sigma^2\right) = o\left(n^{-1}\right),$$

so V_n^1 is a first-order unbiased estimator.

In general, there could be many choices of $\mathbf{B}$ that satisfy (4) and (5), thus giving rise to many first-order unbiased estimators. Among these we want to identify one with minimal variance. Let us approximate the variance of the estimator using the limiting distribution given in Theorem 2; that is,

$$Var\left(\overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}_n\right) \approx tr(\mathbf{CBCB}),$$

which is a convex function of $\mathbf{B}$ (here tr denotes the trace of a matrix). Thus we are led to a convex optimization problem: We want to choose $\mathbf{B}$ to minimize

$$tr(\mathbf{CBCB})$$

subject to the constraints (4) and (5). The first-order optimality conditions are a set of k^2+2 linear equations. Thus we can find the matrix $\mathbf{B}$ that minimizes the asymptotic variance subject to the bias constraints by solving a system of k^2+2 linear equations. We let V_n^1 denote the estimator defined by (8) with $\mathbf{B}$ chosen to solve the constrained optimization problem.

The construction of a second-order unbiased estimator is analogous to the construction we gave for the first-order unbiased estimator. In this case we add the constraints (6) and (7). If $\mathbf{B}$ satisfies these additional constraints and we define

$$V_n^2 = \frac{\overline{\mathbf{W}}_n^T \mathbf{B} \overline{\mathbf{W}}}{1 + \beta_1/n + \beta_2/n^2},$$

then Theorem 4 implies that

$$E\left(V_n^2 - \sigma^2\right) = o\left(n^{-2}\right);$$

that is, V_n^2 is a second-order unbiased estimator. Choosing $\mathbf{B}$ to minimize the asymptotic variance now requires solving a system of $k^2 + 4$ linear equations.

One can also choose $\mathbf{B}$ to satisfy only (4) and minimize the asymptotic variance. We denote this estimator by V_n^0.

5 OVERLAPPING BATCHES

Given a simulation output of length n, $Y_1, Y_2, \ldots, Y_n$, and a "batch size" b, $0 < b < n$, we can construct estimators based on the simulation output $Y_{j+1}, Y_{j+2}, \ldots, Y_{j+b}$ for $0 \leq j \leq n-b$ and then average the estimators. This approach has been used in the construction of several variance constant estimators; see Meketon and Schmeiser (1984) and Alexopoulos et al. (2006).

For $0 < b < n$ and $0 \leq j \leq n-b$ consider the estimator for the jth "batch" of the form

$$V_{j,n} = \widetilde{\mathbf{W}}_{j,j+b}^T \mathbf{B} \widetilde{\mathbf{W}}_{j,j+b},$$

where $\widetilde{\mathbf{W}}_{j,j+b} = \left(\widetilde{W}_{j,j+b}^0, \ldots, \widetilde{W}_{j,j+b}^k\right)$,

$$\widetilde{W}_{j,j+b}^0 = \sum_{i=j+1}^{j+b} Y_i,$$

and for $r \geq 1$,

$$\widetilde{W}_{j,j+b}^r = r \sum_{i=j+1}^{n} \widetilde{W}_{m,i}^{r-1}.$$

The following Lemma shows how to obtain the vectors $\widetilde{\mathbf{W}}_{j,j+b}$ from the vectors $\widetilde{\mathbf{W}}_{0,j+b} = \widetilde{\mathbf{W}}_{j+b}$ produced by the basic simulation algorithm described in Section 3

Lemma 1.

$$\widetilde{\mathbf{W}}_{j,j+b} = \widetilde{\mathbf{W}}_{j+b} - \mathbf{D}\widetilde{\mathbf{W}}_j, \tag{9}$$

where

$$D(r,q) = \begin{cases} \binom{r}{r-q} b^{\overline{r-q}}, & q \leq r, \\ 0, & q > r \end{cases}$$

and $a^{\overline{r}}$ *is the rising factorial function defined by*

$$a^{\overline{r}} = (a)(a+1)(a+2)\cdots(a+r-1)$$

and $a^{\overline{0}} = 1$*; see Knuth (1997).*

Now

$$\overline{\mathbf{W}}_{j,j+b} = (\mathbf{I} - \mathbf{R})(\mathbf{A}\mathbf{N}_b)^{-1} \widetilde{\mathbf{W}}_{j,j+b}$$

is the analog of the vector $\overline{\mathbf{W}}_n$ on which we based the estimators in Section 3, but defined for the jth batch of size b. Then

$$V_{j,n} = \left(\widetilde{\mathbf{W}}_{j+b} - \mathbf{D}\widetilde{\mathbf{W}}_j\right)^T \mathbf{B}_b \left(\widetilde{\mathbf{W}}_{j+b} - \mathbf{D}\widetilde{\mathbf{W}}_j\right),$$

where

$$\mathbf{B}_b = \left((\mathbf{I} - \mathbf{R})\mathbf{N}_b^{-1}\mathbf{A}^{-1}\right)^T \mathbf{B} \left((\mathbf{I} - \mathbf{R})\mathbf{N}_b^{-1}\mathbf{A}^{-1}\right).$$

Our overlapped estimator is then

$$V_n^{ov} = \frac{1}{n-b+1} \sum_{j=0}^{n-b} V_{j,n}.$$

By choosing $\mathbf{B}$ as described in Section 4 we can obtain first- or second-order unbiased estimators. We need to replace n with b, so for example the first-order unbiased estimator has bias $o(b^{-1})$ as $b \to \infty$.

6 NUMERICAL EXPERIMENTS

The numerical results are for a first-order autoregressive process defined by

$$Y_i = \varphi Y_{i-1} + \varepsilon_i, \quad i \geq 1,$$

with the $\varepsilon_i \sim N(0,1)$ independent and $Y_0 \sim N(0,1)$. We set the parameter $\varphi = 0.9$, which results in $\sigma^2 = 19$.

In all cases we constructed the overlapping estimators based on a batch size of $b = n/20$. Figures 1, 2, and 3 show the results of experiments for run lengths of $2,000$, $6,000$, and $10,000$, respectively. The curves for each of the mth order unbiased estimators start at the upper left for $k = 2$ and as k increases the bias generally increases while the variance decreases.

For each run length, the figures show the sample bias and sample variance of the standard overlapping batch means estimator (Alexopoulos et al. 2006), and the first and second order unbiased estimators. The choice of k ranged from 2 to 4. Each experiment consisted of 10^5 independent replications.

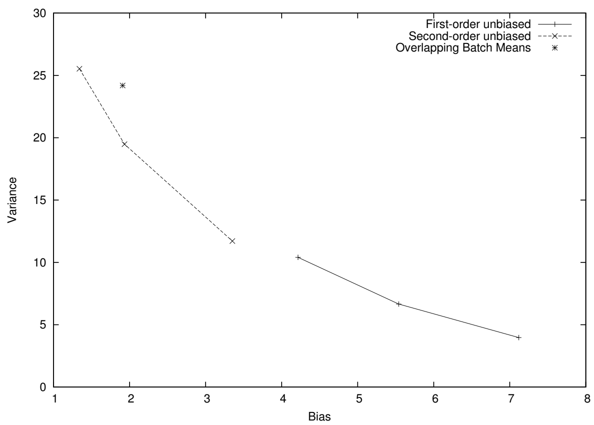

Figure 1: Sample bias and variance for AR(1) simulations, varying k, $n = 2,000$.

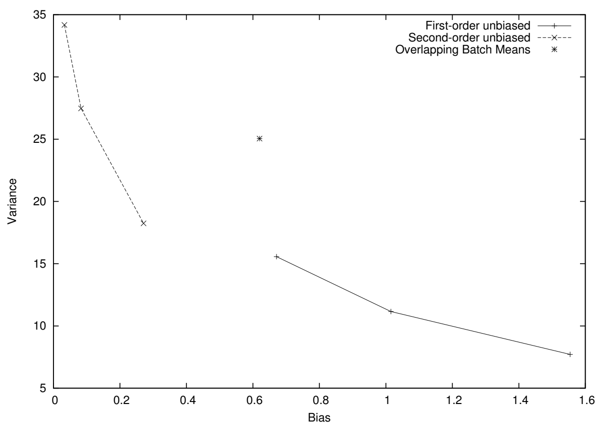

Figure 2: Sample bias and variance for AR(1) simulations, varying k, $n = 6,000$.

As the run length n increases, the bias of the 2nd order unbiased estimator becomes very small compared to that of the batch means estimator.

Experiments with the consistent estimator V_n^0 resulted, as expected, in large bias and low variance, and were omitted from the figures.

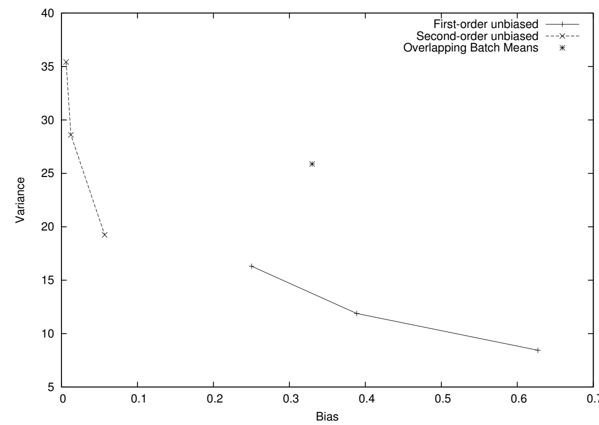

Figure 3: Sample bias and variance for AR(1) simulations, varying k, $n = 10,000$.

REFERENCES

Alexopoulos, C., N. Argon, D. Goldsman, N. Steiger, G. Tokol, and J. Wilson. 2006. Efficient computation of overlapping variance estimators for simulation. *To appear in INFORMS Journal on Computing*.

Calvin, J. M. 2007. Simulation output analysis using integrated paths. To appear in ACM Transactions on Modeling and Computer SImulation; available via `<http://web.njit.edu/∼calvin/ipe.pdf>` [accessed June 10, 2007].

Knuth, D. 1997. *The art of computer programming, volume 1: Fundamental algorithms*. third ed. Reading, Massachusetts: Addison-Wesley.

Meketon, M. S., and B. W. Schmeiser. 1984. Overlapping batch means: something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, 227–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHY

JAMES M. CALVIN is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University and is an associate editor for *ACM Transactions on Modeling and Computer Simulation* and *INFORMS Journal on Computing*. In addition to simulation output analysis, his research interests include global optimization and probabilistic analysis of algorithms.