Bias-Aware Linear Combinations of Variance Estimators

Wenchi Chiu
Department of Industrial Engineering
and Engineering Management
National Tsing Hua University
Hsinchu 300, Taiwan
wenchichiu@gmail.com

David Goldsman
School of Industrial
and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
sman@gatech.edu

Wheyming T. Song
Department of Industrial Engineering
and Engineering Management
National Tsing Hua University
Hsinchu 300, Taiwan
wheyming@ie.nthu.edu.tw

Abstract: A prototype problem in the analysis of steady-state stochastic processes is that of estimating the variance of the sample mean. A commonly used performance criterion for variance estimators is the mean-squared-error (mse) — the sum of the variance and the squared bias. In this paper, we attempt to minimize the variance of an estimator subject to a bias constraint — a goal that differs from that of minimizing mse, in which case there would be no explicit bias constraint. We propose a *bias-aware* mechanism to achieve our goal. Specifically, we use linear combinations of estimators based on different batch sizes to approximately satisfy the bias constraint; and then we minimize the variance by choosing appropriate linear combination weights. We illustrate the use of this mechanism by presenting bias-aware linear combinations of several variance estimators, including non-overlapping batch means, overlapping batch means, and standardized time series weighted area estimators. We also evaluate our mechanism with Monte Carlo examples.

Key–Words: Simulation, Mean-squared-error, Variance Estimation, Non-overlapping Batch Means, Overlapping Batch Means, Standardized Time Series, Weighted Area Estimator.

1 Introduction

One of the classical problems in the statistical analysis of a steady-state autocorrelated stochastic process concerns the estimation of process performance measures, for instance, the expected delay time of a packet in a network system or the expected bit error rate in a communications system [27, 9].

When undertaking the estimation of a performance measure, two issues ought to be addressed. First, how should one obtain a good estimator of the performance measure? Second, how should one evaluate the quality of such an estimator? The research in this paper is motivated by these issues.

To put things on a formal footing, consider a covariance stationary stochastic process $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ with unknown population mean μ, unknown marginal variance R_0, and unknown weighted sums of correlations $\gamma_j \equiv \sum_{h=-\infty}^{\infty} |h|^j \rho_h$, $j = 0, 1$, where ρ_h denotes the lag-h correlation of the process, $\mathrm{Corr}(Y_i, Y_{i+h})$, $h = 0, \pm 1, \pm 2, \ldots$. The value of μ is typically estimated by the sample mean, $\bar{Y}_n \equiv \sum_{i=1}^{n} Y_i/n$; and the variance of sample mean, $\mathrm{Var}(\bar{Y}_n)$, is the quality measure of the sample mean, though some papers use the related *variance parameter*, $\sigma^2 \equiv \lim_{n\to\infty} n\mathrm{Var}(\bar{Y}_n)$, where, under fairly general conditions (e.g., Corollary 2 of [1]), $\sigma^2 = \gamma_0 R_0$ and

$$\mathrm{Var}(\bar{Y}_n) = \frac{\sigma^2}{n} - \frac{\gamma_1 R_0}{n^2} + o(\frac{1}{n^2}).$$

As a practical example, suppose that Y_i is the delay time for packet i at some node in a network. The sample average delay time for the node, $\bar{Y}_n$, is used to estimate the population's mean delay time, and the variance of the average delay time, Var($\bar{Y}_n$), is an indicator of the precision of the average delay time. However, since the delay time data typically have an unknown correlation structure — even if we model this type of problem as, say, a Markovian queueing model [7, 10] — the quantity Var($\bar{Y}_n$) will most likely be unknown; and the problem of providing a good estimator for Var($\bar{Y}_n$) is difficult. In fact, the goal of this paper lies in estimating $\mathrm{Var}(\bar{Y}_n)$.

Various variance estimators have been proposed in the simulation literature to estimate $\mathrm{Var}(\bar{Y}_n)$ [6, 17]. As will be explained in Section 2, many of the popular estimators incorporate batching, e.g., non-overlapping batch means (NBM) [8, 20], overlapping batch means (OBM) [18], and standardized time series area (STS.A) [22] estimators. In addition, it turns out that many estimators are in fact linear combinations of other estimators. For example, OBM can be viewed as a linear combination of certain NBM estimators.

In order to evaluate and compare different variance estimators, a proper statistical analysis dictates that one should take into account the three performance criteria of bias, variance, and mean-squared-error (mse) — the sum of the variance and squared bias. Although mse is a function of both bias and variance, considering mse alone is sometimes not enough. For example, an estimator with a small mse may have relatively high bias yet low variance, leading the user to be "very confident about the wrong answer." For simplicity, we use the term *bias-hazard* to refer to a scenario in which a small mse is accompanied by what we regard as an unacceptably large bias. Due to the possibility of bias-hazard, it is appropriate to consider bias and variance separately.

As techniques from this paper such as batching and linear combinations are applied, the bias-hazard sometimes becomes more severe. For instance, as discussed in Section 2, the use of batching typically decreases estimator variance at the expense of increased estimator bias. Alternatively, one could decrease estimator bias by constructing an appropriate linear combination of basic estimators; but then care must be taken to avoid a variance increase. In this paper, our approach is to control the bias and variance by adjusting estimator operating parameters such as batch sizes and linear-combination weights. Based on this mechanism, we present a bias-aware linear combination of estimators that approximately bounds the bias at a user-defined value and then minimizes the variance.

The rest of the paper is organized as follows. In Section 2, we provide background, including several useful variance estimators and a discussion on linear combinations of the variance estimators. Section 3 proposes the bias-aware mechanism and demonstrates its use through a generic linear-combination estimator. To evaluate the mechanism empirically, Monte Carlo simulation examples are carried out in Section 4. Section 5 concludes the paper. See [2] for a related formulation.

2 Background

In this section, we present background material on various batching variance estimators, including the NBM, OBM, and STS.A estimators and their linear combinations. Throughout the following discussion, we make the reasonable assumption [1] that

$$\left.\begin{array}{c}\boldsymbol{Y} \text{ is covariance stationary with } |\rho_h| = O(\delta^h) \\ \text{for } h = 1, 2, \ldots, \text{ where } \delta \in (0,1).\end{array}\right\} \quad (1)$$

2.1 Non-Overlapping Batch Means

The NBM estimator, first discussed in [5] and [8], transforms correlated data into a few approximately independent non-overlapped batch means. Specifically, the NBM method divides the n observations into b contiguous, non-overlapping batches, each of which contains m consecutive observations. The NBM estimator for $\mathrm{Var}(\bar{Y}_n)$ with batch size m is defined as

$$\hat{V}^{\mathrm{N}}(m) \equiv \frac{m}{n(b-1)} \sum_{i=1}^{b} (\bar{Y}_{i,m} - \bar{Y}_n)^2,$$

where $\bar{Y}_{i,m} \equiv \sum_{j=1}^{m} Y_{(i-1)m+j}/m$ is the ith batch mean, $i = 1, 2, \ldots, b$, and $\bar{Y}_n$ is the overall sample mean. Under Assumption (1), the expected value of $\hat{V}^{\mathrm{N}}(m)$ [1, 13] is

$$\mathrm{E}[\hat{V}^{\mathrm{N}}(m)] = \mathrm{Var}(\bar{Y}_n) - \frac{\gamma_1 R_0}{mn} + o(\frac{1}{mn}). \quad (2)$$

Under additional mild conditions, the asymptotic variance of the NBM estimator [26] is

$$\lim_{\substack{m\to\infty \\ n/m\to\infty}} \frac{n^3}{m} \mathrm{Var}[\hat{V}^{\mathrm{N}}(m)] = 2\sigma^4. \quad (3)$$

2.2 Overlapping Batch Means

The OBM estimator from Meketon and Schmeiser [18] is a weighted average of NBM estimators. It divides the n observations into $n-m+1$ overlapping batches, each consisting of m consecutive observations; of course, the sample means from these batches are highly correlated since they contain common observations. The OBM estimator for $\mathrm{Var}(\bar{Y}_n)$ with batch size m is defined as

$$\hat{V}^{\mathrm{O}}(m) \equiv \frac{m}{(n-m)(n-m+1)} \sum_{i=1}^{n-m+1} (\bar{Y}^{\mathrm{O}}_{i,m} - \bar{Y}_n)^2,$$

where $\bar{Y}^{\mathrm{O}}_{i,m} \equiv \sum_{j=0}^{m-1} Y_{i+j}/m$, $i = 1, 2, \ldots, n-m+1$. Under Assumption (1), the expected value of the OBM estimator [1, 13] is

$$\mathrm{E}[\hat{V}^{\mathrm{O}}(m)] = \mathrm{Var}(\bar{Y}_n) - \frac{\gamma_1 R_0}{mn} + o(\frac{1}{mn}). \quad (4)$$

Under additional mild conditions, the asymptotic variance [26] is

$$\lim_{\substack{m\to\infty \\ n/m\to\infty}} \frac{n^3}{m} \mathrm{Var}[\hat{V}^{\mathrm{O}}(m)] = \frac{4}{3}\sigma^4. \quad (5)$$

2.3 Batched Weighted Area Estimator

The standardized time series weighted area (STS.A) estimator for $\mathrm{Var}(\bar{Y}_n)$, proposed by Schruben [22], uses a functional central limit theorem to transform the process $\boldsymbol{Y}$ into a process that is asymptotically distributed as a Brownian bridge [4]. Goldsman and Schruben [15] proposed the batched version of the STS.A estimator, which is defined as

$$\hat{V}^{\mathrm{A}}(f;m) \equiv \frac{1}{b}\sum_{i=1}^{b}\hat{V}_i^{\mathrm{A}}(f;m),$$

where the weighted area estimator computed from batch i is defined as

$$\hat{V}_i^{\mathrm{A}}(f;m) \equiv \left[\frac{1}{m\sqrt{n}}\sum_{k=1}^{m} f(\frac{k}{m})\sigma T_{i,m}(\frac{k}{m})\right]^2$$

for $i = 1, 2, \ldots, b$.

In addition, the standardized time series from batch i, $T_{i,m}(t)$, is defined as

$$T_{i,m}(t) \equiv \frac{\lfloor mt \rfloor(\bar{Y}_{i,m} - \bar{Y}_{i,\lfloor mt \rfloor})}{\sigma\sqrt{m}}$$

for $0 \le t \le 1$ and $i = 1, 2, \ldots, b$, where $\lfloor\cdot\rfloor$ is the greatest integer function and $\bar{Y}_{i,j} \equiv \sum_{k=1}^{j} Y_{(i-1)m+k}/j$ for $i = 1, 2, \ldots, b$, and $j = 1, 2, \ldots, m$. The weighting function $f(t)$ is continuous on $[0, 1]$, and is chosen to satisfy $\int_0^1\int_0^1 f(s)f(t)(\min(s,t) - st)\,ds\,dt = 1$ (a normalizing constraint). In the current paper we shall exclusively use the weighting function $f_0(t) \equiv \sqrt{12}$, which was the original weight proposed in [22]. Other viable choices of $f(t)$ can be found in [11, 14, 16]. Under Assumption (1), the expected value of STS.A [11, 14] is

$$\mathrm{E}[\hat{V}^{\mathrm{A}}(f;m)] = \mathrm{Var}(\bar{Y}_n) - \frac{[(F(1) - \bar{F}(1))^2 + \bar{F}(1)^2]\gamma_1 R_0}{2mn} + o(\frac{1}{mn}), \tag{6}$$

where $F(t) \equiv \int_0^t f(s)\,ds$ and $\bar{F}(t) \equiv \int_0^t F(s)\,ds$, $0 \le t \le 1$. Under additional mild conditions, the asymptotic variance of $\hat{V}^{\mathrm{A}}(f;m)$ [26] is

$$\lim_{\substack{m\to\infty \\ n/m\to\infty}} \frac{n^3}{m}\mathrm{Var}[\hat{V}^{\mathrm{A}}(f;m)] = 2\sigma^4. \tag{7}$$

2.4 Linear Combinations of Variance Estimators

In this section, we review the idea of using linear combinations of variance estimators to produce a better estimator for $\mathrm{Var}(\bar{Y}_n)$. We refer to the estimators used to form the linear combinations as *component* estimators. Suppose the generic variance estimators $\hat{V}(m_1), \hat{V}(m_2), \ldots, \hat{V}(m_w)$ are selected as the component estimators, where the m_i's represent different batch sizes and *all estimators use the same overall sample size* n. For simplicity, we henceforth assume that the $\hat{V}$'s are either all $\hat{V}^{\mathrm{N}}$'s, all $\hat{V}^{\mathrm{O}}$'s or all $\hat{V}^{\mathrm{A}}$'s. A linear combination $\hat{V}^{\mathrm{C}}$ of these component estimators with coefficients $c_1, c_2, \ldots, c_w$ is

$$\hat{V}^{\mathrm{C}} \equiv \sum_{i=1}^{w} c_i \hat{V}(m_i).$$

In terms of better statistical properties (e.g., small bias, variance, and mse), the estimator $\hat{V}^{\mathrm{C}}$ could provide better performance than any individual component estimator for estimating $\mathrm{Var}(\bar{Y}_n)$ — at least if we choose appropriate weights, $c_1, c_2, \ldots, c_w$. For example, OBM estimators can be viewed as a linear combination of many NBM estimators [18]. Other examples in the context of OBM and standardized time series include [3, 12, 19, 22].

Song and Schmeiser [24] derived the mse-optimal-linear combination weights c_1 and c_2 for two component estimators $\hat{V}(m_1)$ and $\hat{V}(m_2)$ as follows:

$$c_1 \equiv \frac{\mathrm{Var}(\bar{Y}_n)(e_1\tau_2^2 - e_2\tau_{12})}{e_1^2\tau_2^2 + e_2^2\tau_1^2 + \tau_1^2\tau_2^2 - 2e_1e_2\tau_{12} - \tau_{12}^2},$$

and

$$c_2 \equiv \frac{\mathrm{Var}(\bar{Y}_n)(e_2\tau_1^2 - e_1\tau_{12})}{e_1^2\tau_2^2 + e_2^2\tau_1^2 + \tau_1^2\tau_2^2 - 2e_1e_2\tau_{12} - \tau_{12}^2},$$

where $e_i \equiv \mathrm{E}[\hat{V}(m_i)]$ and $\tau_i^2 \equiv \mathrm{Var}[\hat{V}(m_i)]$, for $i = 1, 2$, and $\tau_{12} \equiv \mathrm{Cov}[\hat{V}(m_1), \hat{V}(m_2)]$. Due to the fact that these quantities are unknown and difficult to estimate, the mse-optimal-linear combination weights proposed by Song and Schmeiser [24] can not easily be implemented in practice.

3 Bias-Aware Mechanism

The goal of this paper is to find the optimal linear combination of parameterized variance estimators in terms of minimizing variance, subject to a constraint on the magnitude of the bias. The simplest case ($w = 2$) is to choose $\hat{V}(m_1)$ and $\hat{V}(m_2)$ with batch sizes m_1 and m_2 and coefficients c and $(1 - c)$ to form the linear-combination estimator

$$\hat{V}^{\mathrm{C}} \equiv c\hat{V}(m_1) + (1 - c)\hat{V}(m_2)$$

such that $\hat{V}^{\mathrm{C}}$ satisfies

$$\begin{array}{ll} \text{minimize} & \mathrm{Var}(\hat{V}^{\mathrm{C}}) \\ \text{subject to:} & |\,\mathrm{Bias}(\hat{V}^{\mathrm{C}})\,| < a, \end{array} \tag{8}$$

where a is a constant determined by the user. Practically, the value a could be set to be relatively small compared with $\hat{V}^{\mathrm{C}}$, say $0.1\mathrm{E}[\hat{V}^{\mathrm{C}}]$ or $0.05\mathrm{E}[\hat{V}^{\mathrm{C}}]$. Of course, $\mathrm{E}[\hat{V}^{\mathrm{C}}]$ will likely be unknown in practice, but the user may skirt the issue by conducting a preliminary Monte Carlo study to estimate the quantity.

3.1 Determination of Weights

We discuss properties of linear-combination estimators with $w = 2$, parameterized by batch sizes m and $\lfloor rm \rfloor$, with $0 < r < 1$. First of all, consider the bias results for estimators $\hat{V}^{\mathrm{N}}$, $\hat{V}^{\mathrm{O}}$, and $\hat{V}^{\mathrm{A}}$, stated in Equations (2), (4), and (6), respectively. We will show that the bias can be reduced by choosing certain linear-combination coefficients c and $(1-c)$ so that the $O(1/(mn))$ bias term vanishes; and consequently the bias is reduced substantially when the sample size n is large. The optimal coefficients for the NBM, OBM, and STS.A estimators are stated below (see also [1, 3, 12]).

Theorem 1. *Let $\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)$ be the linear combination of two estimators $\hat{V}(m)$ and $\hat{V}(\lfloor rm \rfloor)$ (both using the same overall sample size n) with coefficients c and $1-c$, where $c = 1/(1-r)$ and $0 < r < 1$. That is,*

$$\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor) \equiv \frac{\hat{V}(m)}{1-r} - \frac{r\hat{V}(\lfloor rm \rfloor)}{1-r}.$$

The expected value of $\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)$ is

$$\mathrm{E}[\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)] = \mathrm{Var}(\bar{Y}_n) + o(\frac{1}{mn}). \qquad (9)$$

Equation (9) for $\hat{V} = \hat{V}^{\mathrm{N}}$, $\hat{V}^{\mathrm{O}}$, and $\hat{V}^{\mathrm{A}}$ can be derived based on the results in Equations (2), (4), and (6), respectively.

3.2 Determination of Parameters

We can argue heuristically that certain values of m tend to result in lower bias. For instance, from Equation (9), we see that if the total number of observations n is fixed, then $|\mathrm{Bias}[\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)]| = o(1/(mn))$ tends to decrease as a function of the batch size m — at least for those variance estimators under study herein.

Meanwhile, Equations (3), (5), and (7) imply that for the variance estimators under consideration, $\mathrm{Var}[\hat{V}(m)]$ tends to increase as the batch size m increases (with fixed n) — which makes sense since larger m corresponds to a smaller number of batches n/m. Thus, the first two terms in the following expression for the variance of the linear-combination estimator,

$$\begin{aligned} &\mathrm{Var}[\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)] \\ &= \frac{\mathrm{Var}[\hat{V}(m)]}{(1-r)^2} + \frac{r^2\mathrm{Var}[\hat{V}(\lfloor rm \rfloor)]}{(1-r)^2} \\ &\quad - \frac{2r\mathrm{Cov}[\hat{V}(m), \hat{V}(\lfloor rm \rfloor)]}{(1-r)^2}, \end{aligned} \qquad (10)$$

tend to increase as the batch size m increases; and so we might expect that $\mathrm{Var}[\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)]$ will increase in m (with fixed n) as well. Equation (10) as well as the work in Goldsman et al. [12] also suggest that the variance of the linear-combination estimator tends to increase as r becomes larger, $0 < r < 1$.

Based on the above heuristic discussion, we will attempt to find the smallest estimator parameters m and r subject to the batch size m being large enough to satisfy the bias constraint. We describe the parameter determination algorithm as follows. In the first step, we initialize the values of r and m; in particular, we set $r = 0.5$ (as described in Goldsman et al. [12]) and estimate m using an existing optimal-mse batch-size estimator (e.g., from Song [23]). Then, we can obtain $\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)$ by combining $\hat{V}(m)$ and $\hat{V}(\lfloor rm \rfloor)$. The next step is to try to decrease the values of r and m while satisfying the bias constraint. We start from the initial parameters r and m, and then decrease m but fix r until the bias constraint is violated. After this, we decrease r, but now fix m until the bias constraint is violated. The process is depicted in Fig. 1. Since we decrease r in fixed decrements, and we decrease m by factors of 2 (and since both r and m are obviously bounded from below by 0 and 1, respectively), the algorithm will eventually stop.

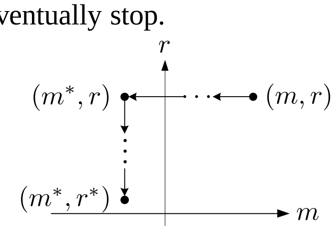

Figure 1: Initial parameters (m, r) and optimal parameters (m^*, r^*)

Parameter Determination Algorithm:

Step 1. Initialization.
  Set $k_r = 0, k_m = 0$, $r = 0.5$, and determine the batch size.

Step 2. Obtain $\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)$, which is a linear combination of $\hat{V}(m)$ and $\hat{V}(\lfloor rm \rfloor)$.

Step 3. Check the validity of the bias constraint.
  If m satisfies the bias-constrained function $\mathrm{Bias}(\hat{V}^{\mathrm{C}}(m, \lfloor rm \rfloor)) < a$, go to (3.1), else go to (3.2).
  (3.1) Check index k_r. If $k_r = 0$, go to (3.1.1), else go to (3.1.2).
    (3.1.1) Set $m^* = m$. Update $m = 0.5m$, $k_m = k_m + 1$, and go to Step 2.
    (3.1.2) Set $r^* = r$. Update $r = r - 0.1$, $k_r = k_r + 1$, and go to Step 2.
  (3.2) Check index k_m. If $k_m = 0$, go to (3.2.1), else go to (3.2.2).

(3.2.1) Update $m = 2m$. If $m \geq n$, stop with no solution, else set $k_m = k_m + 1$ and go to Step 2.

(3.2.2) Check index k_r. If $k_r = 0$, go to (3.1.2), else obtain $\hat{V}^{\text{C}}(m^*, \lfloor r^* m^* \rfloor)$ and stop.

4 Simulation Experiments

In this section we present Monte Carlo examples illustrating the performance of the bias-aware variance estimators. Our examples involve a steady-state first-order autoregressive (AR(1)) process and a symmetric two-state Markov chain (S2MC). The AR(1) process is defined as $Y_{i+1} = \phi Y_i + \epsilon_{i+1}$, $i = 1, 2, \ldots, n$, where the ϵ_i's are independent and identically normal distributed with mean 0 and variance $(1 - \phi^2)R_0$. The S2MC is defined as a two-state dependent symmetric Bernoulli process Y_i, $i = 1, 2, \ldots, n$, with state space $\{d_1, d_2\}$ and transition matrix

$$P \equiv \left[\begin{array}{cc} \frac{1}{2}(1+\rho) & \frac{1}{2}(1-\rho) \\ \frac{1}{2}(1-\rho) & \frac{1}{2}(1+\rho) \end{array} \right],$$

where $|\rho| \leq 1$, $d_1 = \mu - R_0^{1/2}$, and $d_2 = \mu + R_0^{1/2}$.

For both stochastic processes, the sample sizes n considered are 512, 1024, and 2048. Our experiments' parameters are selected as follows: the mean $\mu = 0$; the variance of the sample mean Var$(\bar{Y}_n) = 1$; the sum of correlations $\gamma_0 = \sum_{i=-\infty}^{\infty} \text{Corr}(Y_1, Y_{1+i}) = 10$, which is regarded as a *moderate* correlation structure. (See [25] for more insight on this experimental set up.) We set the bias constraint as $|\text{Bias}(\hat{V})| < 0.1\hat{V}$. In this paper, we test the ideal cases in which $\text{E}[\hat{V}]$ is known for the AR(1) and S2MC cases.

For a particular variance estimator $\hat{V}$ and sample size n, let $m^\star$ denote the batch size that minimizes the mse. Similarly, let $m^\star_B$ denote the variance-optimal batch size under the bias constraint; and let $(m^\star_{c_1}, m^\star_{c_2})$ denote the variance-optimal batch sizes for a linear-combination estimator under the bias constraint.

We examine the empirical performance of $\hat{V}^{\text{N}}(m)$, $\hat{V}^{\text{O}}(m)$, and $\hat{V}^{\text{A}}(f_0, m)$ for batch sizes $m^\star$ and $m^\star_B$, as well as the linear-combination estimators (using the obvious notation) $\hat{V}^{\text{NC}}(m^\star_{c_1}, m^\star_{c_2})$, $\hat{V}^{\text{OC}}(m^\star_{c_1}, m^\star_{c_2})$, and $\hat{V}^{\text{AC}}(f_0; m^\star_{c_1}, m^\star_{c_2})$. The numerical results for optimal batch size, bias, variance, and mse are shown in Tables 1–3. Table 1 summarizes results using the batch size $m^\star$ that minimizes the mse (under no additional constraints). We see from the table that for all of the estimators under consideration, the bias^2 and variance terms are within an order of magnitude of each other.

Table 1: Min-mse Batch Size ($m^\star$) and Performance Results

n	Estimator	Bias	Variance	MSE
		AR(1)		
512	$\hat{V}^{\text{N}}(22)$	0.221	0.058	0.106
	$\hat{V}^{\text{O}}(26)$	0.200	0.061	0.101
	$\hat{V}^{\text{A}}(f_0; 43)$	0.273	0.103	0.177
1024	$\hat{V}^{\text{N}}(27)$	0.167	0.040	0.068
	$\hat{V}^{\text{O}}(32)$	0.158	0.039	0.064
	$\hat{V}^{\text{A}}(f_0; 57)$	0.207	0.077	0.120
2048	$\hat{V}^{\text{N}}(36)$	0.125	0.029	0.044
	$\hat{V}^{\text{O}}(42)$	0.118	0.026	0.040
	$\hat{V}^{\text{A}}(f_0; 75)$	0.163	0.054	0.081
		S2MC		
n	Estimator	Bias	Variance	MSE
512	$\hat{V}^{\text{N}}(25)$	0.173	0.038	0.068
	$\hat{V}^{\text{O}}(25)$	0.195	0.027	0.064
	$\hat{V}^{\text{A}}(f_0; 47)$	0.231	0.081	0.135
1024	$\hat{V}^{\text{N}}(27)$	0.165	0.022	0.049
	$\hat{V}^{\text{O}}(35)$	0.139	0.024	0.043
	$\hat{V}^{\text{A}}(f_0; 54)$	0.221	0.047	0.096
2048	$\hat{V}^{\text{N}}(38)$	0.116	0.019	0.033
	$\hat{V}^{\text{O}}(42)$	0.117	0.015	0.029
	$\hat{V}^{\text{A}}(f_0; 76)$	0.159	0.041	0.066

However, all of the bias values violate the constraint $|\text{Bias}(\hat{V})| < 0.1\text{E}[\hat{V}] = 0.1$. Table 2 gives performance results using variance-optimal batch sizes $(m^\star_B)$, where all of the estimators satisfy the bias constraint; that is, the bias values in the third column are all under 0.1. Comparing Tables 1 and 2, we immediately see that for both the AR(1) and S2MC processes, satisfying the bias constraint is achieved at a substantial increase in the variance and mse. Table 3 gives performance results using variance-optimal batch sizes $(m^\star_{c_1}, m^\star_{c_2})$ under the bias constraint for the linear-combination estimators. Like Table 2, the bias values in the third column in Table 3 are all under 0.1, thus satisfying the bias constraint. Unlike Table 2, almost all of the mse values in Table 3 are also smaller than the corresponding mse values in Table 1 for both the AR(1) and S2MC processes. This shows that the linear-combination estimators preserve comparatively low mse, while satisfying the bias constraint.

Tables 4 and 5 compare results from Tables 2 and 3 (both satisfying the bias constraint) in terms of the variance and mse reductions, respectively. We observe that the variance reduction ranges from 20% to 50% for the AR(1) process and from 20% to 76% for the S2MC process. We also observe the mse reduction ranges from 17% to 39% for the AR(1) process and from 26% to

Table 2: Variance-Optimal Batch Size ($m_B^\star$) and Performance Results Under the Bias Constraint

	AR(1)			
n	Estimator	Bias	Variance	MSE
512	$\hat{V}^{\mathrm{N}}(37)$	0.083	0.143	0.150
	$\hat{V}^{\mathrm{O}}(57)$	0.096	0.163	0.173
	$\hat{V}^{\mathrm{A}}(f_0; 74)$	0.078	0.285	0.292
1024	$\hat{V}^{\mathrm{N}}(41)$	0.087	0.074	0.082
	$\hat{V}^{\mathrm{O}}(52)$	0.099	0.069	0.079
	$\hat{V}^{\mathrm{A}}(f_0; 86)$	0.100	0.150	0.160
2048	$\hat{V}^{\mathrm{N}}(46)$	0.099	0.039	0.049
	$\hat{V}^{\mathrm{O}}(51)$	0.099	0.033	0.042
	$\hat{V}^{\mathrm{A}}(f_0; 108)$	0.088	0.094	0.102
	S2MC			
n	Estimator	Bias	Variance	MSE
512	$\hat{V}^{\mathrm{N}}(37)$	0.077	0.093	0.099
	$\hat{V}^{\mathrm{O}}(52)$	0.095	0.103	0.112
	$\hat{V}^{\mathrm{A}}(f_0; 74)$	0.068	0.222	0.227
1024	$\hat{V}^{\mathrm{N}}(38)$	0.093	0.043	0.052
	$\hat{V}^{\mathrm{O}}(50)$	0.096	0.044	0.053
	$\hat{V}^{\mathrm{A}}(f_0; 86)$	0.099	0.115	0.125
2048	$\hat{V}^{\mathrm{N}}(46)$	0.098	0.027	0.036
	$\hat{V}^{\mathrm{O}}(50)$	0.098	0.021	0.030
	$\hat{V}^{\mathrm{A}}(f_0; 103)$	0.098	0.069	0.078

77% for the S2MC process. In summary, in terms of variance and mse when the bias is bounded, the proposed linear-combination estimator is superior to the existing optimal estimators.

5 Conclusion

Estimation of the mean and variance parameter (or, almost equivalently, the variance of the sample mean) is an important problem in the context of steady-state simulation output analysis. Motivated by the problem that minimizing mse might lead users to be confident about the wrong answer, we have proposed the bias-aware linear-combination variance estimator, along with an algorithm to construct it. Our initial simulation analysis shows that the proposed estimator not only bounds the bias (at least in our idealized examples), but also reduces the variance and mse in many cases.

In order to make the estimator more-useable on real-life examples, we still need to rigorously incorporate into our algorithm better estimators for the bias of a particular estimator (used in Step 3 of our algorithm), as well as better estimators for the optimal batch size (used in the initialization step of the algorithm). These are topics of ongoing study.

Table 3: Variance-Optimal Batch Sizes $(m_{c_1}^\star, m_{c_2}^\star)$ and Performance Results Under the Bias Constraint for the Linear-Combination Estimators

	AR(1)			
n	Estimator	Bias	Variance	MSE
512	$\hat{V}^{\mathrm{NC}}$(32,3)	0.099	0.113	0.123
	$\hat{V}^{\mathrm{OC}}$(32,6)	0.088	0.113	0.117
	$\hat{V}^{\mathrm{AC}}(f_0; 64, 6)$	0.098	0.186	0.209
1024	$\hat{V}^{\mathrm{NC}}$(32,3)	0.092	0.056	0.065
	$\hat{V}^{\mathrm{OC}}$(32,3)	0.093	0.047	0.055
	$\hat{V}^{\mathrm{AC}}(f_0; 64, 12)$	0.091	0.110	0.128
2048	$\hat{V}^{\mathrm{NC}}$(16,8)	0.100	0.019	0.029
	$\hat{V}^{\mathrm{OC}}$(32,3)	0.089	0.023	0.030
	$\hat{V}^{\mathrm{AC}}(f_0; 64, 12)$	0.089	0.060	0.068
	S2MC			
n	Estimator	Bias	Variance	MSE
512	$\hat{V}^{\mathrm{NC}}$(15,7)	0.090	0.035	0.043
	$\hat{V}^{\mathrm{OC}}$(15,7)	0.096	0.024	0.033
	$\hat{V}^{\mathrm{AC}}(f_0; 35, 17)$	0.082	0.152	0.159
1024	$\hat{V}^{\mathrm{NC}}$(15,7)	0.092	0.017	0.025
	$\hat{V}^{\mathrm{OC}}$(15,7)	0.096	0.011	0.021
	$\hat{V}^{\mathrm{AC}}(f_0; 38, 17)$	0.091	0.096	0.105
2048	$\hat{V}^{\mathrm{NC}}$(15,7)	0.090	0.008	0.016
	$\hat{V}^{\mathrm{OC}}$(15,7)	0.095	0.006	0.014
	$\hat{V}^{\mathrm{AC}}(f_0; 41, 20)$	0.092	0.049	0.057

Table 4: Variance Reduction Using Linear-Combination Estimators (Table 3 vs. Table 2)

	AR(1)			S2MC		
m	512	1024	2048	512	1024	2048
NBM	20%	24%	50%	62%	61%	69%
OBM	30%	32%	29%	76%	73%	73%
STS.A	34%	20%	36%	69%	73%	20%

Table 5: Mse Reduction Using Linear-Combination Estimators (Table 3 vs. Table 2)

	AR(1)			S2MC		
m	512	1024	2048	512	1024	2048
NBM	17%	20%	39%	56%	51%	55%
OBM	32%	30%	27%	70%	61%	52%
STS.A	28%	20%	33%	33%	16%	26%

References:

[1] T. Aktaran-Kalaycı, C. Alexopoulos, D. Goldsman, and J. R. Wilson (2006). Exact expected values of variance estimators for simulation. To appear in Naval Research Logistics.

[2] T. Aktaran-Kalaycı, D. Goldsman, and J. R. Wilson (2007). Linear combinations of overlapping

variance estimators for simulation, Operations Research Letters 35, 439–447.

[3] D. Batur, D. Goldsman, and S.-H. Kim (2006). An improved standardized time series Durbin-Watson variance estimator for steady-state simulation, Technical Report, School of ISyE, Georgia Tech, Atlanta, GA.

[4] P. Billingsley (1968). Convergence of Probability Measures, John Wiley & Sons, New York.

[5] R. B. Blackman and J. W. Tukey (1959). The measurement of power spectra from the point of view of communications engineering, Dover Publications Inc., New York (originally published in Bell System Tech. Journal 37 (1958), 185–282 and 485–569).

[6] P. Bratley, B. L. Fox, and L. E. Schrage (1987). A Guide to Simulation, second ed., Springer-Verlag, New York.

[7] I. Cidon, R. Guerin, A. Khamisy, and M. Sidi (1993). Analysis of a correlated queue in a communication system, Proceedings of INFOCOM'93, 209–216.

[8] R. W. Conway, B. W. Johnson, and W. L. Maxwell (1959). Some tactical problems of digital systems simulation, Management Science 6, 92–110.

[9] N. Cvejic, T. Seppäen, and D. Bull (2005). Audio watermark detection improvement by using noise modelling, Proceedings of the 5th WSEAS International Conference on Multimedia, Internet and Video Technologies, 160–164.

[10] K. W. Fendick and V. R. Saksena (1989). Dependence in packet queues, IEEE Transactions on Communications 37, 1173–1183.

[11] R. D. Foley and D. Goldsman (1999). Confidence intervals using orthonormally weighted standardized time series, ACM Transactions on Modeling and Computer Simulation 15, 1–16.

[12] D. Goldsman, K. Kang, S.-H. Kim, A. F. Seila, and G. Tokol (2006). Combining standardized time series area and Cramér-von Mises variance estimators. To appear in Naval Research Logistics.

[13] D. Goldsman and M. S. Meketon (1986). A comparison of several variance estimators, Technical Report J-85-12, School of ISyE, Georgia Tech, Atlanta, GA.

[14] D. Goldsman, M. S. Meketon, and L. W. Schruben (1990). Properties of standardized time series wieghted area variance estimators, Management Science 36, 602–612.

[15] D. Goldsman, and L. W. Schruben (1984). Asymptotic properties of some confidence interval estimators for simulation output, Management Science 30, 1217–1225.

[16] D. Goldsman, and L. W. Schruben (1990). New confidence interval estimators using standardized time series, Management Science 36, 393–397.

[17] A. W. Law and W. D. Kelton (2000). Simulation Modeling and Analysis, third ed., McGraw-Hill, New York.

[18] M. S. Meketon and B. W. Schmeiser (1984). Overlapping batch means: Something for nothing?, Proceedings of the 1984 Winter simulation Conference, 227–230.

[19] A. C. Pedrosa and B. W. Schmeiser (1994). Estimating variance of the sample mean: Optimal batch-size estimation and 1-2-1 overlapping batch means, Technical Report SMS94-3, School of IE, Purdue University, W. Lafayette, IN.

[20] B. W. Schmeiser (1982). Batch-size effects in the analysis of simulation output, Operations Research 30, 556–568.

[21] B. W. Schmeiser and W. T. Song (1987). Correlation among estimators of the variance of the sample mean, Proceedings of the 1987 Winter simulation Conference, 309–317.

[22] L. W. Schruben (1983). Confidence interval estimation using standardized time series, Operations Research 31, 1090–1108.

[23] W. T. Song (1996). On the estimation of optimal batch sizes in the analysis of simulation output, European Journal of Operational Research 88, 304–319.

[24] W. T. Song and B. W. Schmeiser (1988). Minimum-mse linear combinations of variance estimators of the sample mean, Proceedings of the 1988 Winter Simulation Conference, 414–421.

[25] W. T. Song and B. W. Schmeiser (1988). Estimating standard errors: Empirical behavior of asymptotic mse-optimal batch sizes, Computing Science and Statistics: Proceedings of the Symposium on the Interface, 575–580.

[26] W. T. Song and B. W. Schmeiser (1995). Optimal mean-squared-error batch sizes, Management Science 41, 110–123.

[27] D. Tapia-Sanchez, M. Nakano-Miyatake, and H. Perez-Meana (2007). A parallel form constant module algorithm for blind multi-user detection in a multicarrier CDMA receiver, Proceedings of the 6th WSEAS International Conference on Signal Processing, 134–139.