Large-Sample Results for Batch Means

Chiahon Chien David Goldsman Benjamin Melamed

May 12, 1996

Abstract

In analyzing the output process generated by a steady-state simulation, we often seek to estimate the expected value of the output. The sample mean based on a finite sample of size n is usually the estimator of choice for the steady-state mean; and a measure of the sample mean's precision is the variance parameter, i.e., the limiting value of the sample size multiplied by the variance of the sample mean as n becomes large. This paper establishes asymptotic properties of the conventional batch-means (BM) estimator of the variance parameter as both the batch size and the number of batches become large. In particular, we show that the BM variance estimator is asymptotically unbiased and convergent in mean square. We also provide asymptotic expressions for the variance of the BM variance estimator. Exact and empirical examples illustrate our findings.

Authors' addresses: Chiahon Chien, Synopsys, Inc., Mountainview, CA 94043, U.S.A., chien@Synopsys.com; David Goldsman, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, U.S.A., sman@isye.gatech.edu; Benjamin Melamed, Department of Management Science and Information Systems, School of Business – New Brunswick, Rutgers University and RUTCOR, New Brunswick, NJ 08903, U.S.A., melamed@rutcor.rutgers.edu.

Keywords: Simulation, Stationary Process, Variance Estimation, Batch Means Estimator.

1 Introduction

The problem of estimating the mean μ of a stationary process $\{X_i\}_{i=1}^{\infty}$ often arises in the context of steady-state computer simulation. For example, what is the limiting mean transit time of a product going through a complicated manufacturing system? Or what is the long-term expected cost per period incurred by a certain inventory policy?

The obvious point estimator for μ is the sample mean based on n observations, $\bar{X}_n \equiv \sum_{i=1}^{n} X_i/n$. But such a point estimator is not enough to constitute a proper analysis, for it is a random variable (r.v.). Thus, it is also prudent to estimate $\sigma_n^2 \equiv n\mathsf{Var}[\bar{X}_n]$, since this provides a measure of the sample mean's precision. An estimator for σ_n^2 can also be used as a basis for a confidence interval for μ, and for comparing the means of competing systems.

A quantity closely related to σ_n^2 is the so-called *variance parameter*, $\sigma^2 \equiv \sum_{-\infty}^{\infty} R_i$, where $R_i \equiv \mathsf{Cov}[X_1, X_{1+i}]$ $(i \geq 1)$ is the autocovariance function. Under mild conditions (see Lemma 6 below), $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$; so it is of interest to estimate σ^2 as well as (or instead of) σ_n^2.

The literature contains many estimators for σ_n^2 and σ^2 (see Bratley, Fox, and Schrage 1987). This paper studies the popular *batch means* (BM) estimator for σ^2, defined as

$$\hat{V}_B(b,m) \;\equiv\; \frac{m}{b-1}\sum_{i=1}^{b}(\bar{X}_{i,m} - \bar{X}_n)^2 \;=\; \frac{m}{b-1}\Big(\sum_{i=1}^{b}\bar{X}_{i,m}^2 - b\bar{X}_n^2\Big),$$

where we divide the n observations into b adjacent, nonoverlapping batches, each of size m (assume $n = mb$), and $\bar{X}_{i,m} \equiv \sum_{k=1}^{m} X_{(i-1)m+k}/m$ $(1 \leq i \leq b)$. The estimator's appellation is motivated by the fact that we refer to $\bar{X}_{i,m}$ as the ith *batch mean*. For large m, the experimenter assumes that the batch means are approximately independent and identically distributed (i.i.d.) normal r.v.'s with mean μ and unknown variance; this assumption justifies estimation of $\sigma^2/m \approx \sigma_m^2/m = \mathsf{Var}[\bar{X}_{i,m}]$ by the sample variance of the batch means.

The BM estimator has been widely studied in the case that b is *fixed* and $m \to \infty$ (see, e.g., Glynn and Whitt 1991 and Schmeiser 1982). In the case of fixed b, mild conditions yield $\hat{V}_B(b,m) \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b-1)/(b-1)$ as $m \to \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $\chi^2(\nu)$ is the chi-squared distribution with ν degrees of freedom. Assuming uniform integrability (see Billingsley 1968), we also find that $\mathsf{E}[\hat{V}_B(b,m)] \to \sigma^2$ and $\mathsf{Var}[\hat{V}_B(b,m)] \to 2\sigma^4/(b-1)$ as $m \to \infty$. Notice that for fixed b, the BM estimator is not consistent, as $\mathsf{Var}[\hat{V}_B(b,m)]$ does not necessarily go to zero as the batch size m increases. Although point and confidence estimators based on BM with fixed b remain valid, Glynn and Iglehart (1990) show that these estimators are inferior to those based on consistent estimation of σ^2.

The goal of this paper is to see what happens when *both* the batch size $m \to \infty$ *and* the number of batches $b \to \infty$. In §2, we list some preliminary definitions and results that will be invoked afterwards. In §3, we use simple moment and mixing conditions to establish a (well-known) expression for $\mathsf{E}[\hat{V}_B(b,m)]$ and then our main finding that $\lim_{m\to\infty,b\to\infty} b\,\mathsf{Var}[\hat{V}_B(b,m)] = 2\sigma^4$. These results yield expressions for the mean squared error (MSE) of $\hat{V}_B(b,m)$ as an estimator of σ^2. In particular, we show that if both $m \to \infty$ and $b \to \infty$, then $\mathsf{MSE}[\hat{V}_B(b,m)] \to 0$. So $\hat{V}_B(b,m)$ converges in mean square to σ^2 as m and $b \to \infty$. We also give suggestions on how to choose the "optimal" batch size in order to minimize $\mathsf{MSE}[\hat{V}_B(b,m)]$ for fixed n. Exact and empirical examples in §4 illustrate our findings, and §5 provides a brief discussion.

A number of other papers have also attacked the case in which both $m \to \infty$ and $b \to \infty$. For instance, Chien (1989) gives a number of order results on various moments of r.v.'s related to the BM variance estimator; he uses these results to prove the validity of Edgeworth-expansion-based confidence intervals for the mean of a stationary process. Carlstein (1986) proves the mean-square consistency of $\hat{V}_B(b,m)$ as an estimator for σ^2; he assumes uniform integrability of certain r.v.'s whereas we give explicit moments and mixing requirements on the underlying stochastic process. Damerdji (1993) uses strong approximation techniques to establish the mean-square consistency of $\hat{V}_B(b,m)$; his assumptions on the underlying stochastic process are also different than those invoked in the current paper. Song and Schmeiser (1995) obtain expressions for $\mathsf{E}[\hat{V}_B(b,m)]$ in terms of the autocovariance function R_i and argue that $\mathsf{Var}[\hat{V}_B(b,m)]$ takes on a certain explicit form; they use these results to find the batch size that (asymptotically) minimizes $\mathsf{MSE}[\hat{V}_B(b,m)]$.

2 Preliminaries

In this section, we state some definitions and useful results.

Definitions. Suppose $\{X_i\}_{i=-\infty}^{\infty}$ is a stationary process. Following Billingsley (1986), we say that a *σ-field* is a class of subsets of a set that is closed under complements and countable unions. The σ-field *generated* by $\{X_i\}$ is the smallest σ-field with respect to which every X_i is measurable. Let $\mathcal{M}_j^k$ $(j \le k)$ denote the σ-field generated by $\{X_i\}_{i=j}^k$. We say that $\{X_i\}$ is *ϕ-mixing* if for all $j \ge 1$ and $k \ge 1$, and any events $A \in \mathcal{M}_{-\infty}^j$ and $B \in \mathcal{M}_{j+k}^{\infty}$, we have $|P(B|A) - P(B)| \le \phi_k$, where $\phi_k \to 0$ and, without loss of generality, the ϕ_k are nonincreasing. Informally, ϕ-mixing means that events in the distant future are approximately independent of those in the past. Lastly, the notation $f(k) = O(g(k))$ means that $|f(k)/g(k)| \le C$ for some constant C and all $k \ge 1$, and $f(k) = o(g(k))$ means that $f(k)/g(k) \to 0$ as $k \to \infty$.

We will have occasion to use certain tools that are listed here for easy reference. We start off with some simple moment inequalities.

Lemma 1 (Cauchy-Schwarz inequality; see Rohatgi 1976, p. 165.) If $\mathsf{E}[X^2] < \infty$ and $\mathsf{E}[Y^2] < \infty$, then $|\mathsf{E}[XY]| \le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]}$ and $|\mathsf{Cov}[X,Y]| \le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]}$.

Lemma 2 If $\mathsf{E}[X^4]$, $\mathsf{E}[Y^2]$, and $\mathsf{E}[Z^2]$ are all finite, then $|\mathsf{Cov}[X^2, YZ]| \leq \mathsf{E}^{1/2}[X^4]\mathsf{E}^{1/4}[Y^4]\mathsf{E}^{1/4}[Z^4]$.

Lemma 3 (Billingsley 1968, Lemma 1, p. 170.) If X is measurable $\mathcal{M}_{-\infty}^{j}$, Y is measurable $\mathcal{M}_{j+k}^{\infty}$ $(k \geq 1)$, $\mathsf{E}[X^2] < \infty$, and $\mathsf{E}[Y^2] < \infty$, then $|\mathsf{Cov}[X, Y]| \leq 2\sqrt{\phi_k \mathsf{E}[X^2]\mathsf{E}[Y^2]}$.

Lemma 4 (Lyapunov inequality; see Rohatgi 1976, p. 103.) If $\mathsf{E}[|X|^{\ell}] < \infty$ and $1 \leq j \leq k \leq \ell$, then $\mathsf{E}^{1/j}[|X|^j] \leq \mathsf{E}^{1/k}[|X|^k]$.

Lemma 5 (Chien 1989, Proposition 3.1(d).) If $\{X_i\}$ is stationary and $k \geq 1$, then $\mathsf{E}[|X_1|^k] < \infty$ implies that $\mathsf{E}[|\bar{X}_n|^j] < \infty$ for all $1 \leq j \leq k$.

We will sometimes need to have finer moment information than Lemma 5's crude finiteness results. Lemmas 6 and 7 below require the quantity $\gamma \equiv -2\sum_{j=1}^{\infty} jR_j$ (see Song and Schmeiser 1995). Note that γ is well-defined if $\mathsf{E}[X_i^2] < \infty$ and $\phi_k = O(k^{-4-\epsilon})$ for some $\epsilon > 0$, for then Lemma 3 implies

$$|\gamma| \;\leq\; 2\sum_{j=1}^{\infty} j|\mathsf{Cov}(X_1, X_{1+j})| \;\leq\; 4\mathsf{E}[X_1]\sum_{j=1}^{\infty} j\phi_j^{1/2} \;<\; \infty.$$

Lemma 6 (Song and Schmeiser 1995; Titus 1985, p. 29.) If $\{X_i\}$ is stationary, $\mathsf{E}[X_1^4] < \infty$, and $\phi_k = O(k^{-4-\epsilon})$ for some $\epsilon > 0$, then $\mathsf{Var}[\bar{X}_n] = \sigma^2/n + \gamma/n^2 + o(1/n^2)$.

Lemma 7 (Chien 1989, Corollary 2.9 and Proposition 3.1). If $\{X_i\}$ is stationary, $\mathsf{E}[X_1] = 0$, $\mathsf{E}[X_1^{12}] < \infty$, and $\phi_k = O(k^{-8-\epsilon})$ for some $\epsilon > 0$, then $\mathsf{Var}[\bar{X}_n^2] = 2\sigma^4/n^2 + O(1/n^3)$ and $\mathsf{E}[\bar{X}_n^4] = 3\sigma^4/n^2 + O(1/n^3) < \infty$.

Remark 1 We can justify Lemma 7 heuristically. By a central limit theorem, $\bar{X}_n/\sqrt{\mathsf{Var}[\bar{X}_n]} \xrightarrow{\mathcal{D}} \mathrm{Nor}(0,1)$ as $n \to \infty$. Then the continuous mapping theorem (see Billingsley 1968, Theorem 5.1) implies that $n\bar{X}_n^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(1)$. Assuming uniform integrability and taking variances, we get $n^2\,\mathsf{Var}[\bar{X}_n^2] \to 2\sigma^4$.

The next lemma follows from the fact that $\sum_{i=1}^{\infty} i^{-\beta}$ is finite for $\beta > 1$.

Lemma 8 If $\phi_k = O(k^{-\alpha})$ for $\alpha > 2$, then $\sum_{i=1}^{\infty} \phi_{im}^{1/2} = O(m^{-\alpha/2})$.

The following lemma will be used in the proof of the main variance expression in §3.

Lemma 9 (Chien 1989, Lemma A.10(1).) If $\{X_i\}$ is stationary, $\mathsf{E}[X_1] = 0$, $\mathsf{E}[X_1^{12}] < \infty$, and $\phi_k = O(k^{-9})$, then $\mathsf{Cov}[\sum_{i=1}^{b} \bar{X}_{i,m}^2, \bar{X}_n^2] = O(1/b)O(1/m^2)$.

Remark 2 We provide a different proof of Lemma 9 in the Appendix. The alternative proof is carried out under slightly more restrictive conditions than those of Chien (1989), but it is somewhat simpler and makes the current paper self-contained.

To complete the series of preliminary lemmas, we state one last result that will be invoked in an analytical example in §4.1 involving a first-order moving average process.

Lemma 10 (See, e.g., Patel and Read 1982.) If (X, Y) is a bivariate normal r.v. with $\mathsf{E}[X] = \mathsf{E}[Y] = 0$, then $\mathsf{Cov}[X^2, Y^2] = 2\mathsf{Cov}^2[X, Y]$.

With the requirements of the lemmas in mind, we shall henceforth adopt the following standing assumptions.

Assumption 1 The process $\{X_i\}$ is stationary and ϕ-mixing with $\mathsf{E}[X_1^{12}] < \infty$ and $\phi_k = O(k^{-9})$. Without loss of generality, we take $\mathsf{E}[X_1] = 0$.

Remark 3 Although the ϕ-mixing assumption has been made mainly for mathematical convenience, we note that many stochastic processes are ϕ-mixing, e.g., ℓ-dependent processes and finite-state Markov chains (see Billingsley 1968, p. 167). Further, in light of Lemma 5, we can henceforth assume that $\mathsf{E}[|\bar{X}_n|^j] < \infty$ for all $1 \le j \le 12$.

3 Results

This section contains our main results on the expected value, variance, and mean squared error of the BM variance estimator $\hat{V}_B(b, m)$.

Lemma 6 and some algebra yield the next well-known result which says that as $m \to \infty$, the BM estimator $\hat{V}_B(b, m)$ becomes unbiased for σ^2 (whether or not $b \to \infty$).

Theorem 1 (Goldsman and Meketon 1986 and Song and Schmeiser 1995.) Under Assumption 1,

$$\mathsf{E}[\hat{V}_B(b, m)] = \frac{m}{b-1}\mathsf{E}\left[\sum_{i=1}^{b} \bar{X}_{i,m}^2 - b\bar{X}_n^2\right] = \frac{mb}{b-1}(\mathsf{Var}[\bar{X}_{i,m}] - \mathsf{Var}[\bar{X}_n]) \tag{1}$$

$$= \sigma^2 + \frac{\gamma(b+1)}{mb} + o(1/m). \tag{2}$$

Our goal now is to obtain an expression for $\mathsf{Var}[\hat{V}_B(b, m)]$ as both m and b become large.

Theorem 2 Under Assumption 1, as $m \to \infty$ and $b \to \infty$,

$$b\,\mathsf{Var}[\hat{V}_B(b,m)] \;=\; \frac{2\sigma^4 b(b+1)}{(b-1)^2} + O(m^{-1/4}) + O(1/b) \;=\; 2\sigma^4 + o(1). \tag{3}$$

Proof We will break up $b\,\mathsf{Var}[\hat{V}_B(b,m)]$ into five terms and treat each separately.

$$\begin{aligned}
b\,\mathsf{Var}[\hat{V}_B(b,m)] &= \frac{m^2 b}{(b-1)^2}\left[\mathsf{Var}\Big[\sum_{i=1}^{b} \bar{X}_{i,m}^2\Big] + b^2\,\mathsf{Var}[\bar{X}_n^2] - 2b\,\mathsf{Cov}\Big[\sum_{i=1}^{b} \bar{X}_{i,m}^2, \bar{X}_n^2\Big]\right] \\
&= \frac{m^2 b}{(b-1)^2}\Bigg[b\,\mathsf{Var}[\bar{X}_{1,m}^2] + b^2\,\mathsf{Var}[\bar{X}_n^2] + 2\sum_{i=2}^{b-1}(b-i)\mathsf{Cov}[\bar{X}_{1,m}^2, \bar{X}_{i+1,m}^2] \\
&\qquad +2(b-1)\mathsf{Cov}[\bar{X}_{1,m}^2, \bar{X}_{2,m}^2] - 2b\,\mathsf{Cov}\Big[\sum_{i=1}^{b} \bar{X}_{i,m}^2, \bar{X}_n^2\Big]\Bigg].
\end{aligned} \tag{4}$$

By Lemma 7, the first two terms on the right side of (4) are

$$\frac{m^2 b^2}{(b-1)^2}\left\{\frac{2\sigma^4}{m^2} + O(\frac{1}{m^3}) + b\left[\frac{2\sigma^4}{n^2} + O(\frac{1}{n^3})\right]\right\} \;=\; \frac{2\sigma^4 b(b+1)}{(b-1)^2} + O(m^{-1}). \tag{5}$$

The third term of (4) is bounded by

$$\begin{aligned}
\frac{2m^2 b}{(b-1)^2}\sum_{i=2}^{b-1}(b-i)|\mathsf{Cov}[\bar{X}_{1,m}^2, \bar{X}_{i+1,m}^2]| &\le 8m^2 \sum_{i=2}^{b-1} |\mathsf{Cov}[\bar{X}_{1,m}^2, \bar{X}_{i+1,m}^2]| \\
&\le 16m^2\,\mathsf{E}[\bar{X}_m^4]\sum_{i=2}^{\infty} 2\sqrt{\phi_{(i-1)m}} \quad \text{(by Lemma 3)} \\
&= O(m^{-9/2}) \quad \text{(by Lemmas 7 and 8 and the assumption that } \phi_k = O(k^{-9})\text{)}.
\end{aligned} \tag{6}$$

We deal with the fourth term of (4). For ease of exposition, suppose that $\sqrt{m}$ is an integer, else use the "floor" (round-down) function $\lfloor \sqrt{m} \rfloor$. Then

$$\begin{aligned}
\frac{2m^2 b}{b-1}\left|\mathsf{Cov}[\bar{X}_{1,m}^2, \bar{X}_{2,m}^2]\right| &\le 4\left|\mathsf{Cov}\left[\bar{X}_{1,m}^2, \Big(\sum_{j=1}^{\sqrt{m}} X_{j+m} + \sum_{j=\sqrt{m}+1}^{m} X_{j+m}\Big)^2\right]\right| \\
&= 4\left|\mathsf{Cov}\left[\bar{X}_{1,m}^2, \Big(\sum_{j=1}^{\sqrt{m}} X_{j+m}\Big)^2 + 2\Big(\sum_{j=1}^{\sqrt{m}} X_{j+m}\Big)\Big(\sum_{j=\sqrt{m}+1}^{m} X_{j+m}\Big) + \Big(\sum_{j=\sqrt{m}+1}^{m} X_{j+m}\Big)^2\right]\right| \\
&\le 8\,\mathsf{E}^{1/2}[\bar{X}_m^4]\left[m\mathsf{E}^{1/2}[\bar{X}_{\sqrt{m}}^4] + m^{3/2}\Big(\mathsf{E}[\bar{X}_{\sqrt{m}}^4]\mathsf{E}[\bar{X}_{m-\sqrt{m}}^4]\Big)^{1/4} + m^2\phi_{\sqrt{m}}^{1/2}\mathsf{E}^{1/2}[\bar{X}_{m-\sqrt{m}}^4]\right] \\
&\qquad \text{(by Lemmas 1, 2, and 3)} \\
&= O(m^{-1/2}) + O(m^{-1/4}) + O(m^{-9/4}) \;=\; O(m^{-1/4}) \\
&\qquad \text{(by Lemma 7 and the assumption that } \phi_k = O(k^{-9})\text{)}
\end{aligned} \tag{7}$$

Finally, Lemma 9 shows that the fifth term of (4) is $O(1/b)$. The theorem follows from this and (5)–(7). $\square$

The next result gives the mean squared error of $\hat{V}_B(b,m)$ as an estimator of σ^2, i.e., $\mathsf{MSE}[\hat{V}_B(b,m)] \equiv (\mathsf{E}[\hat{V}_B(b,m)] - \sigma^2)^2 + \mathsf{Var}[\hat{V}_B(b,m)]$, the squared bias plus the variance of $\hat{V}_B(b,m)$.

Corollary 1 Under Assumption 1,

$$\mathsf{MSE}[\hat{V}_B(b,m)] \;=\; \frac{\gamma^2}{m^2} + \frac{2\sigma^4}{b} + o(1/m^2) + o(1/b).$$

Proof Immediate from (2) and (3). $\square$

Let us consider the problem of selecting m (and b) as a function of n. Suppose we take $m = cn^{\zeta}$, where c is some positive constant and $0 < \zeta < 1$. Then it is easy to see that $\zeta = 1/3$ is asymptotically optimal in the sense that $\mathsf{MSE}[\hat{V}_B(b,m)]$ will have the fastest convergence rate to zero. If we ignore the small-order terms and take $m = cn^{1/3}$, we can minimize

$$\widehat{\mathsf{MSE}}[\hat{V}_B(b,m)] \;\equiv\; \frac{\gamma^2}{m^2} + \frac{2\sigma^4}{b} \;=\; \frac{1}{n^{2/3}}\left[\frac{\gamma^2}{c^2} + 2c\sigma^4\right]$$

with respect to c. We then obtain the approximate "optimal" batch size $\widehat{m}^{\star} \equiv (\gamma^2 n/\sigma^4)^{1/3}$, with resulting approximate "optimal" mean squared error $\widehat{\mathsf{MSE}}^{\star}(n) \equiv 3(\gamma\sigma^4/n)^{2/3}$ (cf. Goldsman and Meketon 1986 and Song and Schmeiser 1995).

4 Examples

This section illustrates our findings. In §4.1, we analytically calculate $\mathsf{E}[\hat{V}_B(b,m)]$, $\mathsf{Var}[\hat{V}_B(b,m)]$, and $\mathsf{MSE}[\hat{V}_B(b,m)]$ for the first-order moving average [MA(1)] process. In §4.2, we present corresponding empirical results for the TES$^+$ process of Melamed (1991).

4.1 Analytical Example

Consider the stationary MA(1) process, $X_j = \theta\epsilon_{j-1} + \epsilon_j$ $(j \geq 1)$, where the ϵ_j are i.i.d. $\mathrm{Nor}(0,1)$ r.v.'s. This is a simple yet interesting process, used widely in time-series applications. The coefficient θ determines the autocovariance structure of the MA(1); in particular,

$$\mathsf{Cov}[X_j, X_{j+k}] \;=\; \begin{cases} 1+\theta^2 & \text{for } k=0 \\ \theta & \text{for } k=\pm 1, \\ 0 & \text{otherwise} \end{cases}$$

so that $\sigma^2 = \sum_{j=-\infty}^{\infty} R_j = (1+\theta)^2$ and $\gamma = -2\sum_{j=1}^{\infty} jR_j = -2\theta$. It is also easy to show that all of the requirements of Assumption 1 hold. Further, we can establish that

$$m^2\mathsf{Cov}[\bar{X}_{1,m}, \bar{X}_{i+1,m}] \;=\; \begin{cases} m\sigma^2+\gamma & \text{for } i=0 \\ -\gamma/2 & \text{for } i=1. \\ 0 & \text{for } i>1 \end{cases} \tag{8}$$

Then (1) and (8) imply that

$$\mathsf{E}[\hat{V}_B(b,m)] \;=\; \frac{mb}{b-1}(\mathsf{Var}[\bar{X}_{i,m}] - \mathsf{Var}[\bar{X}_n]) \;=\; \sigma^2 + \frac{\gamma(b+1)}{mb}, \tag{9}$$

which is in agreement with Theorem 1. Of special interest is the fact that if $\theta \approx -1$, then $\sigma^2 \ll \gamma$; so $\hat{V}_B(b,m)$ is quite biased for small m.

In order to obtain an exact expression for $\mathsf{Var}[\hat{V}_B(b,m)]$, observe that (8) implies

$$bm^2\,\mathsf{Cov}[\bar{X}_{i,m}, \bar{X}_n] \;=\; m^2\sum_{j=1}^{b}\mathsf{Cov}[\bar{X}_{i,m}, \bar{X}_{j,m}] \;=\; \begin{cases} m\sigma^2+\gamma/2 & \text{for } i=1 \text{ or } b \\ m\sigma^2 & \text{for } 2\le i\le b-1. \end{cases} \tag{10}$$

By Lemma 10, we can rewrite (4) as

$$\begin{aligned} b\,\mathsf{Var}[\hat{V}_B(b,m)] \;&=\; \frac{2m^2b}{(b-1)^2}\Big[b\,\mathsf{Var}^2[\bar{X}_m] + b^2\,\mathsf{Var}^2[\bar{X}_n] \\ &\qquad +2\sum_{i=1}^{b-1}(b-i)\mathsf{Cov}^2[\bar{X}_{1,m}, \bar{X}_{i+1,m}] - 2b\sum_{i=1}^{b}\mathsf{Cov}^2[\bar{X}_{i,m}, \bar{X}_n]\Big] \\ &=\; \frac{2m^2b}{(b-1)^2}\left\{b\left[\frac{\sigma^2}{m}+\frac{\gamma}{m^2}\right]^2 + b^2\left[\frac{\sigma^2}{n}+\frac{\gamma}{n^2}\right]^2 + \frac{(b-1)\gamma^2}{2m^4} - \frac{2}{bm^2}\left[b\sigma^4 + \frac{2\gamma\sigma^2}{m} + \frac{\gamma^2}{2m^2}\right]\right\} \\ &\qquad \text{(by (8), (10), and algebra)} \\ &=\; \frac{2\sigma^4 b}{b-1} + \frac{4\gamma\sigma^2(b+1)}{m(b-1)} + \frac{\gamma^2(3b^2-b-2+\frac{2}{b})}{m^2(b-1)^2} \;=\; 2\sigma^4 + o(1) \end{aligned} \tag{11}$$

as $m\to\infty$ and $b\to\infty$, which is in agreement with Theorem 2. By (9) and (11), we also have

$$\mathsf{MSE}[\hat{V}_B(b,m)] \;=\; \frac{2\sigma^4}{b-1} + \frac{\gamma^2(b^4+3b^3-3b^2-2b+3)}{n^2(b-1)^2} + \frac{4\gamma\sigma^2(b+1)}{n(b-1)}, \tag{12}$$

Table 1: Performance of Approximations $\widehat{m}^\star$ and $\widehat{\mathsf{MSE}}^\star(n)$ for MA(1) Process ($\theta = -0.5$).

	Exact Results		Approximations	
n	$m^\star$	$\mathsf{MSE}[\hat{V}_B(b^\star, m^\star)]$	$\widehat{m}^\star$	$\widehat{\mathsf{MSE}}^\star(n)$
100	11	0.04140	11.70	0.02193
200	14	0.02221	14.74	0.01382
400	18	0.01245	18.57	0.00870
800	23	0.00720	23.39	0.00548
1600	29	0.00425	29.47	0.00345
3200	37	0.00256	37.13	0.00218
6400	47	0.00155	46.78	0.00137

in agreement with Corollary 1.

We continue the MA(1) example by studying the performance of the approximations, $\widehat{m}^\star$ and $\widehat{\mathsf{MSE}}^\star(n)$, for the optimal batch size and optimal MSE, respectively. Suppose we take $\theta = -0.5$, so that $\sigma^2 = 0.25$ and $\gamma = 1.0$. Equation (12) allows us to determine the *actual* optimal batch size and optimal $\mathsf{MSE}[\hat{V}_B(b, m)]$, denoted by $m^\star$ and $\mathsf{MSE}[\hat{V}_B(b^\star, m^\star)]$, respectively, where (for ease of exposition) we allow $b^\star \equiv n/m^\star$ to be noninteger. For various n, Table 1 compares the actual $m^\star$ and $\mathsf{MSE}[\hat{V}_B(b^\star, m^\star)]$ with the corresponding approximations. We see that, for this example, $\widehat{m}^\star$ is right on target, while $\widehat{\mathsf{MSE}}^\star(n)/\mathsf{MSE}[\hat{V}_B(b^\star, m^\star)]$ appears to approach unity as n increases. Of course, use of the approximations requires that γ be *known*, a boon that is unlikely to occur in practice; see Song (1993) and Pedrosa and Schmeiser (1994) for procedures to estimate γ.

4.2 Empirical Example

We can study the behavior of $\hat{V}_B(b, m)$ when it is applied to a nontrivial stationary stochastic process—namely, the TES$^+$ process of Melamed (1991). The main merit of TES$^+$ processes is that they can be transformed to fit precisely any prescribed empirical marginal distribution and most empirical autocorrelation structures.

We now describe the TES$^+$ process. Let $\langle x \rangle \equiv x - \lfloor x \rfloor$ be the fractional part of x. Suppose $U_0^+ \sim \mathrm{Unif}(0,1)$, and let $U_i^+ \equiv \langle U_{i-1}^+ + V_i \rangle$ ($i \geq 1$), where $V_1, V_2, \ldots$ is a sequence of i.i.d. r.v.'s such that for each $i \geq 1$, V_i is independent of the past history, $U_0^+, U_1^+, \ldots, U_{i-1}^+$. We denote the common probability density function of the V_i by f_V and that of their j-fold convolution by $f_V^{(j)}$. Then it can be shown that the U_i^+

are correlated, identically distributed (c.i.d.) Unif$(0, 1)$. Next, we define a measurable function $D : [0, 1) \to \mathbf{R}$, used to transform a uniform r.v. into one with a general marginal distribution. Let $D = G^{-1}$ for some cumulative distribution function G. If we set $X_i \equiv D(U_i^+)$ $(i \geq 0)$, then the X_i are c.i.d. with common marginals from G. Jagerman and Melamed (1992a, 1992b) show that the autocovariance function of $\{X_i\}$ is $R_j = 2\sum_{k=1}^{\infty} \widetilde{f}_V^{(j)}(2\pi k\sqrt{-1}\,)|\widetilde{D}(2\pi k\sqrt{-1}\,)|^2$ $(j \geq 0)$, where a tilde denotes the Laplace transform operator.

For a basic TES$^+$ process $\{X_i\}$, we set $V_i \equiv \ell + (r - \ell)W_i$ $(i \geq 1)$, where $\{W_i\}$ is a stream of i.i.d. Unif$(0, 1)$ variates and $-1/2 \leq \ell < r < 1/2$. Then the autocovariance function of the X_i specializes to (see Jagerman and Melamed 1992a, 1992b)

$$R_j \;=\; 2\sum_{k=1}^{\infty}\left(\frac{\sin(\pi k\alpha)}{\pi k\alpha}\right)^j \cos(\pi k\alpha\varphi j)|\widetilde{D}(2\pi k\sqrt{-1}\,)|^2 \;\; (j \geq 0), \tag{13}$$

where $\alpha = r - \ell$ and $\varphi = (r + \ell)/\alpha$. Evaluation of this autocovariance structure is amenable to numerical methods since the power terms decay quickly.

To demonstrate the above ideas, assume that $D(x) = -\ell\text{n}(1 - x)$, so that $X_i \sim$ Exp(1). For this example, we shall set $\varphi = 0.0$ and consider various choices for α. From (13), we find that the R_j decrease monotonically to zero; e.g., when $\alpha = 0.2$, the R_j practically vanish by $j \approx 100$. Further, from (13), we can calculate numerically σ^2, γ, and $\lim_{m\to\infty, b\to\infty} b\,\mathsf{Var}[\hat{V}_B(b, m)] = 2\sigma^4$; for instance, when $\alpha = 0.2$, we have $\sigma^2 = 13.809$, $\gamma = -185.38$, and $2\sigma^4 = 381.38$. To verify these calculations, we ran a small Monte Carlo simulation study for this TES$^+$ process with exponential distortions. Some results of the study are displayed in Table 2. Each entry is based on 2000 independent simulation runs. A $\bar{V}_B$ entry is the sample mean of the 2000 realizations of $\hat{V}_B(b, m)$, and $S^2(\hat{V}_B(b, m))$ is the corresponding sample variance. Indeed, the quantities appear to converge towards the appropriate limiting values.

5 Discussion

The results of §3 and examples of §4 show that the BM variance estimator $\hat{V}_B(b, m)$ behaves as advertised as both the batch size m and number of batches b become large. Specifically, we showed that $\hat{V}_B(b, m)$ is asymptotically unbiased and convergent in mean square. We also provided asymptotic expressions for the variance of $\hat{V}_B(b, m)$.

How can one put our results to practical use? Indeed, a number of interesting BM-based applications arise in situations where the experimenter cannot place an *a priori* upper bound on the number of batches that will be required. For example:

- Fixed-width confidence interval estimation procedures for μ.

Table 2: Performance of $\hat{V}_B(b,m)$ for TES$^+$ Process.

$\alpha = 0.4$				$\alpha = 0.2$			
m	b	$\bar{V}_B$	$bS^2(\hat{V}_B(b,m))$	m	b	$\bar{V}_B$	$bS^2(\hat{V}_B(b,m))$
10	2	2.14	19.66	50	10	9.95	181.65
20	4	2.94	22.13	100	20	11.80	246.30
40	8	3.31	23.66	150	30	12.61	292.15
60	12	3.35	23.51	200	40	12.93	316.63
80	16	3.43	23.78	250	50	13.06	320.15
100	20	3.45	23.95	300	60	13.19	337.48
200	40	3.53	24.78	400	80	13.39	340.21
400	80	3.53	24.97	500	100	13.40	343.75
		$\sigma^2 = 3.57$	$2\sigma^4 = 25.49$			$\sigma^2 = 13.81$	$2\sigma^4 = 381.38$

- Sequential ranking and selection procedures for determining which of a number of steady-state processes has the largest (or smallest) mean.
- Control chart procedures based on sequential probability ratio tests.

These topics are the subjects of ongoing research.

Acknowledgments: We thank David Jagerman, Bruce Schmeiser, Jim Wilson, a referee, and an Associate Editor for helpful suggestions. Our work was supported by National Science Foundation Grant No. DDM-9012020.

Appendix

In this Appendix, we prove a slightly weakened version of Lemma 9, namely:

Lemma 11 If $\{X_i\}$ is stationary, $\mathsf{E}[X_1] = 0$, $\mathsf{E}[X_1^{28}] < \infty$, and $\phi_k = O(k^{-15})$, then $\mathsf{Cov}[\sum_{i=1}^{b} \bar{X}_{i,m}^2, \bar{X}_n^2] = O(1/b)O(1/m^2)$.

The lemma has more stringent moment and mixing requirements than those of Lemma 9; but the new requirements are mild and will not affect the lemma's use in any realistic situation. Before proving Lemma 11, we will need a few preliminary results: Lemma 12 gives two simple covariance relations, Lemma 13 proves a moment inequality, and Lemma 14 concerns moments of the sample mean. (Lemma 14 is responsible for Lemma

11's seemingly restrictive assumptions.) In the ensuing discussion, let $Y_i \equiv \bar{X}_{i,m}$, $\delta_i \equiv \phi^{1/2}_{(i-1)m+1}$ ($i \geq 1$), and $\delta_0 \equiv 1$.

Lemma 12 If $\mathsf{E}[X^8] < \infty$, $\mathsf{E}[Y^4] < \infty$, $\mathsf{E}[Z^4] < \infty$, and $\mathsf{E}[Y] = \mathsf{E}[Z] = 0$, then $\mathsf{Cov}[X^2, YZ] = \mathsf{Cov}[X^2Y, Z] - \mathsf{E}[X^2]\mathsf{Cov}[Y, Z]$ and $\mathsf{Cov}[X^2Y, Z] = \mathsf{Cov}[X^2Z, Y]$.

Lemma 13 If $\{X_i\}$ is stationary and $\mathsf{E}[X_1^8] < \infty$ (so that, by Lemma 5, $\mathsf{E}[Y_1^8] < \infty$), then for $i, j, k \geq 1$, $\mathsf{Cov}[Y_i^2, Y_jY_k] \leq 4\,\mathrm{median}\{\delta_{|i-j|}, \delta_{|j-k|}, \delta_{|i-k|}\}\mathsf{E}^{1/2}[Y_1^8]$.

Proof (of Lemma 13.) By Lemmas 3, 1, and 4, we have

$$\begin{aligned} |\mathsf{Cov}[Y_i^2, Y_jY_k]| &\leq 2\max\{\delta_{|j-i|}, \delta_{|k-i|}\}\sqrt{\mathsf{E}[Y_i^4]\mathsf{E}[Y_j^2Y_k^2]} \\ &\leq 2\max\{\delta_{|j-i|}, \delta_{|k-i|}\}\mathsf{E}[Y_1^4] \leq 2\max\{\delta_{|j-i|}, \delta_{|k-i|}\}\mathsf{E}^{1/2}[Y_1^8]. \end{aligned} \tag{14}$$

By Lemma 12, the triangle inequality, and manipulations similar to those leading to (14),

$$|\mathsf{Cov}[Y_i^2, Y_jY_k]| \leq |\mathsf{Cov}[Y_i^2Y_j, Y_k]| + \mathsf{E}[Y_i^2]|\mathsf{Cov}[Y_j, Y_k]| \leq 4\max\{\delta_{|k-i|}, \delta_{|k-j|}\}\mathsf{E}^{1/2}[Y_1^8]. \tag{15}$$

By symmetry, we have $|\mathsf{Cov}[Y_i^2, Y_jY_k]| \leq 4\max\{\delta_{|j-i|}, \delta_{|k-j|}\}\mathsf{E}^{1/2}[Y_1^8]$, and the result follows from this, (14), and (15). $\square$

Lemma 14 (Chien 1989, Corollary 2.7.) If $\{X_i\}$ is stationary, $\mathsf{E}[X_1] = 0$, $\mathsf{E}[X_1^{4(j-1)}] < \infty$, and $\phi_n = O(n^{-2(j-1+\epsilon)})$ for some $\epsilon > 0$ and $j \geq 1$, then $\mathsf{E}[\bar{X}_n^j] = O(n^{-\lceil j/2 \rceil})$, where $\lceil \cdot \rceil$ is the "ceiling" (round-up) function.

We are now in a position to prove the main result of the Appendix.

Proof (of Lemma 11.) We have

$$\begin{aligned} b^2|\mathsf{Cov}[\sum_{i=1}^b Y_i^2, \bar{X}_n^2]| &\leq \sum_{i=1}^b\sum_{j=1}^b\sum_{k=1}^b |\mathsf{Cov}[Y_i^2, Y_jY_k]| \\ &\leq 4\,\mathsf{E}^{1/2}[Y_1^8]\sum_{i=1}^b\sum_{j=1}^b\sum_{k=1}^b \mathrm{median}\{\delta_{|i-j|}, \delta_{|j-k|}, \delta_{|i-k|}\} \quad \text{(by Lemma 13)} \\ &\leq 24\,\mathsf{E}^{1/2}[Y_1^8]\sum_{i=1}^b\sum_{j=i}^b\sum_{k=j}^b \mathrm{median}\{\delta_{j-i}, \delta_{k-j}, \delta_{k-i}\} \quad \text{(by symmetry)} \\ &\leq 24\,\mathsf{E}^{1/2}[Y_1^8]\sum_{i=1}^b\sum_{j=i}^b\sum_{k=j}^b \min\{\delta_{j-i}, \delta_{k-j}\} \quad (\text{since } \delta_{k-i} \leq \min\{\delta_{j-i}, \delta_{k-j}\}) \end{aligned}$$

$$
\begin{aligned}
&= 24\,\mathsf{E}^{1/2}[Y_1^8]\sum_{i=1}^{b}\sum_{j=0}^{b-i}\sum_{k=0}^{b-i-j}\min\{\delta_j,\delta_k\} \quad \text{(by substituting } j \text{ for } j-i \text{ and } k \text{ for } k-j)\\
&\le 24b\,\mathsf{E}^{1/2}[Y_1^8]\sum_{j=0}^{b}\sum_{k=0}^{b}\min\{\delta_j,\delta_k\}\\
&\le 48b\,\mathsf{E}^{1/2}[Y_1^8]\sum_{j=0}^{b}\sum_{k=0}^{j}\delta_j \quad \text{(by symmetry)}\\
&\le 48b\,\mathsf{E}^{1/2}[Y_1^8]\sum_{j=0}^{b}\delta_j^{1/2}\sum_{k=0}^{j}\delta_k^{1/2} \quad (\text{since } \delta_k \ge \delta_j \text{ when } k \le j)\\
&\le 48b\,\mathsf{E}^{1/2}[Y_1^8]\left(\sum_{k=0}^{\infty}\delta_k^{1/2}\right)^2\\
&= O(b)O(m^{-2}) \quad \text{(by Lemma 14 and assumption).} \quad \square
\end{aligned}
$$

References

P. Billingsley (1968). *Convergence of Probability Measures*. John Wiley and Sons, New York.

P. Billingsley (1986). *Probability and Measure*, 2nd Ed. John Wiley and Sons, New York.

P. Bratley, B. L. Fox, and L. E. Schrage (1987), *A Guide to Simulation*, 2nd Ed. Springer-Verlag, New York.

E. Carlstein (1986). "The Use of Subseries for Estimating the Variance of a General Statistic from a Stationary Sequence." *Annals of Statistics* **14**, 1171–1179.

C.-H. Chien (1989). "Small Sample Theory for Steady State Confidence Intervals." Technical Report No. 37, Department of Operations Research, Stanford University, Stanford, California.

H. Damerdji (1993). "Mean-Square Consistency of the Variance Estimator in Steady-State Simulation Output Analysis." *Operations Research* **43**, 282–291.

P. W. Glynn and D. L. Iglehart (1990). "Simulation Output Analysis Using Standardized Time Series." *Mathematics of Operations Research* **15**, 1–16.

P. W. Glynn and W. Whitt (1991). "Estimating the Asymptotic Variance with Batch Means." *Operations Research Letters* **10**, 431–435.

D. Goldsman and M. S. Meketon (1986). "A Comparison of Several Variance Estimators for Stationary Increment Stochastic Processes." Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

D. L. Jagerman and B. Melamed (1992a). "The Transition and Autocorrelation Structure of TES Processes, Part I: General Theory." *Stochastic Models* **8**, 193–219.

D. L. Jagerman and B. Melamed (1992b). "The Transition and Autocorrelation Structure of TES Processes, Part II: Special Cases." *Stochastic Models* **8**, 499–527.

B. Melamed (1991). "TES: A Class of Methods for Generating Autocorrelated Uniform Variates." *ORSA Journal on Computing* **3**, 317–329.

J. K. Patel and C. B. Read (1982). *Handbook of the Normal Distribution.* Marcel Dekker, New York.

A. Pedrosa and B. W. Schmeiser (1994). "Estimating the Variance of the Sample Mean: Optimal Batch-Size Estimation and 1-2-1 Overlapping Batch Means." Technical Report SMS94-3, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

V. K. Rohatgi (1976). *An Introduction to Probability Theory and Mathematical Statistics.* John Wiley and Sons, New York.

B. W. Schmeiser (1982). "Batch Size Effects in the Analysis of Simulation Output." *Operations Research* **30**, 556–568.

W.-M. T. Song (1993). "On the Estimation of Optimal Batch Sizes in the Analysis of Simulation Output Analysis." Technical Report, Department of Industrial Engineering, National Tsing-Hua University, Republic of China.

W.-M. T. Song and B. W. Schmeiser (1995). "Optimal Mean-Squared-Error Batch Sizes." *Management Science* **41**, 110–123.

B. D. Titus (1985). "Modified Confidence Intervals for the Mean of an Autoregressive Process." Technical Report No. 34, Department of Operations Research, Stanford University, Stanford, California.