Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

TO BATCH OR NOT TO BATCH

Christos Alexopoulos
David Goldsman

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332–0205, U.S.A.

ABSTRACT

When designing steady-state computer simulation experiments, one is often faced with the choice of batching observations in one long run or replicating a number of smaller runs. Both methods are potentially useful in simulation output analysis. In its simplest form, the choice boils down to: Should we divide one long run into b adjacent, nonoverlapping batches, each of size m? Or should we conduct b independent replications, each of length m?

We give results and examples to lend insight as to when one method might be preferred over the other. In the steady-state case, batching and replication perform about the same in terms of estimating the mean and variance parameter, though replication tends to do better than batching when it comes to the performance of confidence intervals for the mean. On the other hand, batching can often do better than replication when it comes to point and confidence-interval estimation of the steady-state mean in the presence of an initial transient. This is not particularly surprising, and is a common rule of thumb in the folklore.

1 INTRODUCTION

The purpose of this article is to compare the methods of batch means and independent replications in the context of steady-state simulation output analysis.

When designing steady-state computer simulation experiments, one is often faced with the choice of batching observations in one long run or replicating a number of smaller runs. Both methods are potentially useful in simulation output analysis, where we might be interested in obtaining confidence intervals (CI's) for the unknown steady-state mean μ, or at least in obtaining estimates for the variance of the sample mean, the obvious point estimator for μ.

In its simplest form, the choice of batching or replicating boils down to: Should we divide one long run into b adjacent, nonoverlapping batches, each of size m? Or should we conduct b independent replications, each only of length m? The trade-offs between the two alternatives are well known: Batching ameliorates the effects of initialization bias (if it is present), but produces batch means that often are correlated; replication yields independent sample means, but may suffer from initialization bias at the beginning of each of the runs. So what should we use for steady-state simulation output analysis — the method of independent replications (IR) or batch means (BM)?

There is a wide literature on the subject, outlined in Alexopoulos and Goldsman (2003). Our analysis complements Whitt (1991), who also studied the problem of one long run versus independent replications based on the efficiency of the estimator of μ.

The organization of the rest of the article is as follows. Section 2 gives some relevant background and notation, while Sections 3 and 4 provide our main findings. Our claims are supported in Section 5, which provides illustrative examples. We end up showing that IR does just fine in the steady-state case, but certain initial transients ruin the performance of IR without straining that of BM too badly. Section 6 wraps up the discussion with some final thoughts.

2 BACKGROUND

In this section we define the problem of interest and present the notation to be used in the sequel.

The goal is to estimate the mean μ of a stationary stochastic process $\{S_\ell,\ \ell \geq 1\}$, e.g., a steady-state simulation output process. The natural point estimator for μ is the sample mean based on n observations, $\bar{S}_n \equiv n^{-1}\sum_{\ell=1}^{n} S_\ell$. A wise, statistically sound practice is to supplement the sample mean with a measure of its precision. Relevant steady-state performance measures are $\sigma_n^2 \equiv n\,\mathsf{Var}(\bar{S}_n)$ and the associated *variance parameter*, $\sigma^2 \equiv \lim_{n\to\infty}\sigma_n^2$. Estimators for σ_n^2 and σ^2 can subsequently be used to obtain CI's for μ, among other things. The two simplest and most widely used approaches for estimating σ_n^2 and σ^2 are IR and BM. For good introductory references on these and other variance estimators, see Alexopoulos and Seila (1998), Fishman (2001), or Law and Kelton (2000).

2.1 Some Notation and Definitions

We use some additional notation and definitions throughout the paper.

The quantities $S_1, S_2, \ldots$ always denote stationary observations. By contrast, $X_1, X_2, \ldots$ denote generic observations — sometimes they will be stationary (Section 3), sometimes not (Section 4). Sometimes the observations will be divided into independent replications, and sometimes into batches. We generally use the notations $\bar{Y}_{1,m}, \bar{Y}_{2,m}, \ldots$ and $\bar{Y}_n$ for the replicate sample means and grand sample mean from all of the replicates, respectively; the analogous notations $\bar{X}_{1,m}, \bar{X}_{2,m}, \ldots$ and $\bar{X}_n$ are reserved for the batch sample means and grand mean from all of the batches.

For a stationary process $\{S_\ell\}$, the autocovariance function is denoted $R_j \equiv \mathsf{Cov}(S_1, S_{1+j}),\ j = 0, \pm 1, \pm 2, \ldots$. This leads to the well-known alternative expressions

$$\sigma_n^2 \;=\; R_0 + 2\sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right)R_i$$

and, if $\sum_{j=1}^{\infty} j|R_j| < \infty$,

$$\sigma^2 \;=\; \sum_{i=-\infty}^{\infty} R_i.$$

In addition, we define the related "center-of-gravity" constant $\gamma \equiv -2\sum_{j=1}^{\infty} jR_j$ (Song and Schmeiser 1995). Along the way, we will also assume that the process is ϕ-mixing (Billingsley 1968). Informally, ϕ-mixing means that events in the distant future are approximately independent of those in the past.

The "little-oh" notation $f(m) = o(g(m))$ means that $f(m)/g(m) \to 0$ as $m \to \infty$. The "big-oh" notation $f(m) = O(g(m))$ means that $|f(m)/g(m)| \le C$ for some constant C and all $m \ge 1$.

2.2 Independent Replications

Here we conduct b independent replications of a simulation process $\{X_\ell\}$. We will assume that the replication length m is fixed and common among the b replications. Denoting by $X_{i,j}$ the jth observation from replication i, we have the following allocation of the $n = bm$ observations.

Replication 1:	$X_{1,1}, X_{1,2}, \ldots, X_{1,m}$
Replication 2:	$X_{2,1}, X_{2,2}, \ldots, X_{2,m}$
$\vdots$	
Replication b:	$X_{b,1}, X_{b,2}, \ldots, X_{b,m}$

For each of these replications, we calculate a replicate (sample) mean $\bar{Y}_{i,m} \equiv m^{-1}\sum_{j=1}^{m} X_{i,j}$, $i = 1, 2, \ldots, b$. By the way in which the replications are run, we see that the replicate means are independent and identically distributed (i.i.d.) random variables (r.v.'s); and if the process $\{X_\ell\}$ is stationary with mean μ, then the replicate means have mean μ and variance $\mathsf{Var}(\bar{Y}_{i,m}) = \sigma_m^2/m$.

The IR estimator for the steady-state mean μ is simply the grand sample mean from the b independent replications, $\bar{Y}_n \equiv b^{-1}\sum_{i=1}^{b} \bar{Y}_{i,m}$. If $\{X_\ell\}$ is stationary with mean μ, then $\mathsf{E}[\bar{Y}_n] = \mu$, and so the grand mean is unbiased for μ.

The IR estimator for σ^2 is

$$\widehat{V}_R \;\equiv\; \frac{m}{b-1}\sum_{i=1}^{b}(\bar{Y}_{i,m} - \bar{Y}_n)^2.$$

Since $\widehat{V}_R$ is m times the sample variance of the replicate means, it follows that $\widehat{V}_R$ is an unbiased estimator for $m\mathsf{Var}(\bar{Y}_{i,m})$, i.e.,

$$\mathsf{E}[\widehat{V}_R] \;=\; m\mathsf{Var}(\bar{Y}_{i,m}). \tag{1}$$

In addition, if $\{X_\ell\}$ is stationary, then $\mathsf{E}[\widehat{V}_R] = \sigma_m^2$.

2.3 Batch Means

Here we conduct one long run of the simulation, say of length n, and we divide the n observations $X_1, X_2, \ldots, X_n$ into b adjacent, nonoverlapping batches, each of size m (assuming that $n = mb$).

Batch 1:	$X_1, X_2, \ldots, X_m$
Batch 2:	$X_{m+1}, X_{m+2}, \ldots, X_{2m}$
$\vdots$	
Batch b:	$X_{(b-1)m+1}, X_{(b-1)m+2}, \ldots, X_n$

For each of these batches, we calculate the *batch mean*, $\bar{X}_{i,m} \equiv m^{-1}\sum_{k=1}^{m} X_{(i-1)m+k}$, for $i = 1, 2, \ldots, b$.

The BM estimator for μ is the grand sample mean from the b batch means, $\bar{X}_n \equiv b^{-1}\sum_{i=1}^{b}\bar{X}_{i,m} = n^{-1}\sum_{\ell=1}^{n} X_\ell$. Stationarity implies that $\mathsf{E}[\bar{X}_n] = \mu$, so the grand mean is unbiased for μ; and the variance of the grand mean is, by definition, $\mathsf{Var}(\bar{X}_n) = \sigma_n^2/n$.

The BM estimator for σ^2 is

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

For large batch size m, the experimenter assumes that the batch means are approximately i.i.d. normal r.v.'s with mean μ and unknown variance $\sigma_m^2/m = \mathsf{Var}(\bar{X}_{i,m})$; this assumption motivates estimation of $\sigma^2 \doteq \sigma_m^2$ by m times the sample variance of the batch means.

3 IR / BM STEADY-STATE COMPARISON

In this section, we assume that the process starts off and remains in steady state. For comparison purposes, we will assume that the IR and BM competitors use equivalent replication/batch size m and numbers of replications/batches b. We will compare the IR and BM estimators as applied to the point estimation of μ (Section 3.1), the point estimation of σ^2 (Section 3.2), and confidence interval estimation for μ (Section 3.3).

3.1 Estimators for the Mean

We compare the mean squared errors (MSE's) of $\bar{Y}_n$ and $\bar{X}_n$, the IR and BM estimators, respectively, for the steady-state mean μ. As pointed out in Sections 2.2 and 2.3, both $\bar{Y}_n$ and $\bar{X}_n$ are unbiased for μ. Thus, $\mathsf{MSE}(\bar{Y}_n;\mu) = \mathsf{Var}(\bar{Y}_n)$ and $\mathsf{MSE}(\bar{X}_n;\mu) = \mathsf{Var}(\bar{X}_n)$. The following lemma provides relevant expressions for these variances.

Lemma 1 (Song and Schmeiser 1995; Titus 1985). If $\{X_\ell\}$ is stationary with $\mathsf{E}[X_1^4] < \infty$, and ϕ-mixing with $\phi_k = O(k^{-4-\epsilon})$ for some $\epsilon > 0$, then

$$\sigma_m^2 \;=\; m\,\mathsf{Var}(\bar{Y}_{i,m}) \;=\; \sigma^2 + \frac{\gamma}{m} + o\left(\frac{1}{m}\right). \qquad (2)$$

Using Lemma 1 and the fact that the replications are independent, we eventually have

$$\mathsf{MSE}(\bar{Y}_n;\mu) - \mathsf{MSE}(\bar{X}_n;\mu) \;=\; \frac{\gamma(b-1)}{n^2} + o\left(\frac{1}{nm}\right),$$

implying that the difference in MSE's is very small.

Remark 1 Most queue waiting processes tend to have a positive autocorrelation structure, for which it turns out that $\gamma < 0$, and σ_n^2 converges to σ^2 from below. So in this case, $\mathsf{MSE}(\bar{Y}_n;\mu)$ is a tad smaller than $\mathsf{MSE}(\bar{X}_n;\mu)$.

3.2 Estimators for the Variance Parameter

We divide the analysis into three portions: Derivations for IR, derivations for BM, and then a comparison.

3.2.1 Independent Replications

Suppose that $\{X_\ell\}$ is stationary and satisfies the conditions of Lemma 1. Then Equations (1) and (2) imply

$$\mathsf{E}[\widehat{V}_R] \;=\; \sigma_m^2 \;=\; \sigma^2 + \frac{\gamma}{m} + o(1/m), \qquad (3)$$

which yields an explicit expression for the bias of $\widehat{V}_R$ as an estimator of σ^2.

Assuming the replicate means have finite and well-defined fourth moments, it can be shown that

$$\begin{aligned} b\,\mathsf{Var}(\widehat{V}_R) &= m^2\,\mathsf{E}[(\bar{Y}_{1,m}-\mu)^4] - \sigma_m^4\left(\frac{b-3}{b-1}\right) \\ &= \frac{2b\sigma^4}{b-1} + O\left(\frac{1}{m}\right) \qquad (4) \end{aligned}$$

(see Kang and Goldsman 1990 and Alexopoulos and Goldsman 2003, among others, for details).

Actually, some more simplification is possible if we are willing to assume that $\bar{Y}_{1,m}$ is normal (e.g., if m is sufficiently large). Then we have the exact result

$$\mathsf{Var}(\widehat{V}_R) \;=\; \frac{2\sigma_m^4}{b-1}, \qquad (5)$$

which looks familiar — for if we assume that the replicate means are i.i.d. normal, then $\widehat{V}_R \sim \sigma_m^2\chi^2(b-1)/(b-1)$, where $\chi^2(\nu)$ is the chi-square distribution with ν degrees of freedom. If we believe this distributional assumption, the variance from Equation (5) follows.

3.2.2 Batch Means

We have some analogous expressions for the expected value and variance of the BM estimator for σ^2, the latter result requiring a couple of additional assumptions on the process $\{X_\ell\}$.

Theorem 1 If the process $\{X_\ell\}$ is stationary with $\sum_{j=1}^{\infty} j|R_j| < \infty$, then

$$\mathsf{E}[\widehat{V}_B] \;=\; \sigma^2 + \frac{\gamma(b+1)}{n} + o\left(\frac{1}{m}\right). \qquad (6)$$

Theorem 2 (Goldsman and Meketon 1986; Song and Schmeiser 1995; and Chien, Goldsman, and Melamed 1997.) Suppose that the process $\{X_\ell\}$ is stationary with $\mathsf{E}[X_1^{12}] < \infty$ and ϕ-mixing with $\phi_k = O(k^{-9})$. Then

$$b\,\mathsf{Var}(\widehat{V}_B) \;=\; 2\sigma^4 + o(1), \qquad (7)$$

the last equality holding as $m \to \infty$ and $b \to \infty$.

In addition, for fixed b, different, but still mild, moment and mixing conditions imply $\widehat{V}_B \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b-1)/(b-1)$ as $m \to \infty$ (cf. Glynn and Whitt 1991).

3.2.3 Comparison

What about the MSE's for the IR and BM estimators for σ^2 in the steady-state case? By Equations (3) and (6), we have

$$\mathsf{Bias}^2(\widehat{V}_R;\sigma^2) - \mathsf{Bias}^2(\widehat{V}_B;\sigma^2) \;=\; -\frac{\gamma^2(2b+1)}{n^2} + o\left(\frac{1}{m^2}\right).$$

By Equations (4) and (7), we have $\mathsf{Var}(\widehat{V}_B) = \mathsf{Var}(\widehat{V}_R) + o(1)$. So, up to the order terms, we cannot really distinguish between the IR and BM variances. And then we see that $\mathsf{MSE}(\widehat{V}_R;\sigma^2)$ and $\mathsf{MSE}(\widehat{V}_B;\sigma^2)$ differ by only small-order terms.

3.3 Confidence Intervals for the Mean

The analysis on CI's for μ turns out to be difficult. On one hand, we assumed in this section that the replicate and batch means all have the same (steady-state) distribution. On the other, we might encounter some problems since neither the replicate nor batch means are necessarily normal, or since the batch means are not even independent. Nevertheless, *for purposes of the rough-cut analysis of the present subsection*, we shall assume for now that the replicate and batch means are identically distributed from the steady-state normal distribution.

3.3.1 Independent Replications

The well-known $100(1-\alpha)\%$ IR CI for μ is

$$\mu \;\in\; \bar{Y}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_R/n}, \qquad (8)$$

where $t_{\beta,\nu}$ is the β quantile of Student's t distribution with ν degrees of freedom. Under the liberal assumptions of this subsection, this CI achieves *perfect coverage*.

Theorem 3 If the replicate means are i.i.d. normal with mean μ, then the probability that the CI (8) will cover μ is exactly $1-\alpha$.

3.3.2 Batch Means

The $100(1-\alpha)\%$ BM CI for μ is

$$\mu \;\in\; \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n}. \qquad (9)$$

Unfortunately, things do not work out as smoothly for the BM CI as they did for the IR CI, even under the liberal assumptions of this subsection, for the batch means are not independent. The good news is that under the mild assumption that the process satisfies a functional central limit theorem, as the batch size m becomes large (with fixed number of batches b), we have

$$\mathsf{Pr}\left(\mu \;\in\; \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n}\right) \;\to\; 1-\alpha$$

(see, e.g., Glynn and Iglehart 1990). Hence, the true coverage probability approaches the nominal value $1-\alpha$. But for small values of m, the estimator $\widehat{V}_B$ is biased for σ^2 (and σ_n^2), and so the coverage probability is often less than nominal for systems with positive autocorrelation (see, e.g., Sargent, Kang, and Goldsman 1992).

Although we cannot give a general expression for the coverage probability, we can at least do so for the special case of $b=2$ batches.

Proposition 1 If $b=2$ batch means $\bar{X}_{1,m}$ and $\bar{X}_{2,m}$ are bivariate normal, both with marginal mean μ and variance σ_m^2/m, then the probability that the CI (9) will cover μ is

$$\begin{aligned}\mathsf{CVG} \;&\equiv\; \mathsf{Pr}\left(\mu \;\in\; \bar{X}_n \pm t_{\alpha/2,1}\sqrt{\widehat{V}_B/n}\right)\\ &=\; 2F_{t(1)}\left(t_{\alpha/2,1}\sqrt{\widetilde{\sigma}_m^2/\sigma_n^2}\right) - 1,\end{aligned}$$

where $F_{t(1)}(\cdot)$ is the cumulative distribution function (c.d.f.) of the $t(1)$ (Cauchy) distribution and

$$\widetilde{\sigma}_m^2 \;\equiv\; \sigma_m^2 - m\mathsf{Cov}(\bar{X}_{1,m},\bar{X}_{2,m}).$$

Example 1 Consider the stationary first-order moving average process, $X_\ell = \epsilon_\ell + \theta\epsilon_{\ell-1}$, $\ell \ge 1$, where the ϵ_ℓ's are i.i.d. Nor(0, 1) r.v.'s. This process has covariance function $R_0 = 1+\theta^2$, $R_{\pm 1} = \theta$, and $R_i = 0$, otherwise.

It is easy to show that

$$\sigma_n^2 \;=\; \sigma^2 + \gamma/n,$$

with

$$\sigma^2 \;=\; (1+\theta)^2 \quad\text{and}\quad \gamma \;=\; -2\theta.$$

Further,

$$\mathsf{Cov}(\bar{X}_{1,m},\bar{X}_{2,m}) \;=\; \theta/m^2.$$

So for the case $n=2m$ ($b=2$), we see that

$$\frac{\widetilde{\sigma}_m^2}{\sigma_n^2} \;=\; \frac{\sigma^2 - \frac{3\theta}{m}}{\sigma^2 - \frac{\theta}{m}}$$

and can conclude from Proposition 1 that the coverage is $<[>]\; 1-\alpha$ if $\theta > [<]\; 0$. $\diamond$

3.4 Steady-State Recap

Almost all of the results in Sections 3.1 and 3.2 indicate that the IR and BM methods perform similarly when it comes to point estimation for μ and σ^2 — except perhaps for the meaningless case of very small sample sizes. However, the results in Section 3.3 seem to say that IR has an advantage over BM in terms of the steady-state performance of the respective CI's for μ. But the victory is hollow, since some serious problems arise if the process under study does not happen to be in steady state.

4 IR / BM TRANSIENT COMPARISON

The main reason for skepticism concerning the use of IR is the stationarity issue; and in this section, we examine what happens in the nonstationary case. Now the observations $\{X_\ell\}$ start off polluted by a transient function, before eventually settling down to steady state. To keep things as simple as possible, and still make our points, we will study the model

$$X_\ell = S_\ell + a_\ell, \tag{10}$$

$\ell = 1, 2, \ldots$, where $\{S_\ell\}$ is a *stationary* process with mean μ, and $\{a_\ell\}$ is simply a sequence of constants converging to zero.

Still more notation. Let $S_{i,j}$ be the jth observation from replication i of the stationary process $\{S_\ell\}$, for $i = 1, 2, \ldots, b$ and $j = 1, 2, \ldots, m$. We denote the replicate means of the process $\{S_\ell\}$ by $\bar{T}_{1,m}, \bar{T}_{2,m}, \ldots, \bar{T}_{b,m}$, i.e., $\bar{T}_{i,m} \equiv m^{-1} \sum_{j=1}^{m} S_{i,j}$, for $i = 1, 2, \ldots, b$. Further, let $\bar{T}_n \equiv b^{-1} \sum_{i=1}^{b} \bar{T}_{i,m}$ be the grand sample mean taken over the b independent replications of the process $\{S_\ell\}$. Thus, defining $\bar{a}_m \equiv m^{-1} \sum_{\ell=1}^{m} a_\ell$, we can express Model (10)'s replicate means in terms of those from the stationary process $\{S_\ell\}$: $\bar{Y}_{i,m} = \bar{T}_{i,m} + \bar{a}_m$, $i = 1, 2, \ldots, b$, Similarly, the grand sample mean of the b replicates of Model (10) is $\bar{Y}_n = \bar{T}_n + \bar{a}_m$. We see from these definitions that the replicate means $\bar{Y}_{1,m}, \bar{Y}_{2,m}, \ldots, \bar{Y}_{b,m}$ under Model (10) are i.i.d. with expected value $\mu + \bar{a}_m$ and variance σ_m^2/m. Further, the grand sample mean $\bar{Y}_n$ of all the replications has expected value $\mu + \bar{a}_m$ and variance σ_m^2/n.

Now we define the analogous notation for the batch means method. To begin with, $S_{(i-1)m+j}$ is the jth observation from batch i of the stationary process $\{S_\ell\}$, for $i = 1, 2, \ldots, b$ and $j = 1, 2, \ldots, m$. We denote the batch means of this process by $\bar{S}_{1,m}, \bar{S}_{2,m}, \ldots, \bar{S}_{b,m}$, i.e., $\bar{S}_{i,m} \equiv m^{-1} \sum_{j=1}^{m} S_{(i-1)m+j}$, for $i = 1, 2, \ldots, b$. Further, let $\bar{S}_n \equiv b^{-1} \sum_{i=1}^{b} \bar{S}_{i,m}$ be the grand sample mean taken over the b batches of $\{S_\ell\}$. Thus, defining $\bar{a}_{i,m} \equiv m^{-1} \sum_{j=1}^{m} a_{(i-1)m+j}$, for $i = 1, 2, \ldots, b$, we can express Model (10)'s batch means in terms of those from the stationary process $\{S_\ell\}$: $\bar{X}_{i,m} = \bar{S}_{i,m} + \bar{a}_{i,m}$, $i = 1, 2, \ldots, b$. Similarly, the grand sample mean of the b batches of Model (10) is $\bar{X}_n = \bar{S}_n + \bar{a}_n$, where $\bar{a}_n \equiv n^{-1} \sum_{\ell=1}^{n} a_\ell$. Thus, the ith batch mean $\bar{X}_{i,m}$ under Model (10) has expected value $\mu + \bar{a}_{i,m}$ and variance σ_m^2/m, for $i = 1, 2, \ldots, b$. Finally, the grand sample mean $\bar{X}_n$ of all the batch means has expected value $\mu + \bar{a}_n$ and variance σ_n^2/n.

With our simple additive transient function $\{a_\ell\}$ in mind, Section 4.1 compares the IR and BM estimators as applied to the point estimation of μ under Model (10), Section 4.2 does the same for σ^2, and Section 4.3 is concerned with the CI estimation for μ.

4.1 Estimators for the Mean

Under Model (10),

$$\mathsf{E}[\bar{Y}_n] = \mathsf{E}[\bar{Y}_{i,m}] = \mu + \bar{a}_m \quad \text{and} \quad \mathsf{E}[\bar{X}_n] = \mu + \bar{a}_n,$$

and (as in Section 3.1)

$$\mathsf{Var}(\bar{Y}_n) = \sigma_m^2/n \quad \text{and} \quad \mathsf{Var}(\bar{X}_n) = \sigma_n^2/n.$$

Then we get

$$\begin{aligned}&\mathsf{MSE}(\bar{Y}_n; \mu) - \mathsf{MSE}(\bar{X}_n; \mu)\\ &\quad = \bar{a}_m^2 - \bar{a}_n^2 + \frac{\gamma(b-1)}{n^2} + o\left(\frac{1}{nm}\right).\end{aligned}$$

As we commented in Section 3.1, the last two terms in this difference are probably very small. Thus, it may very well be that the bulk of the difference in the MSE's is contained in the first term, $\bar{a}_m^2 - \bar{a}_n^2$. If the underlying stochastic process $\{X_\ell\}$ eventually reaches steady state, then we must have $a_\ell \to 0$ as $\ell \to \infty$. So it is reasonable to assume that $\bar{a}_m^2 > \bar{a}_n^2$; and if m is small enough and n is large enough, it may even be the case that $\bar{a}_m^2 \gg \bar{a}_n^2$.

4.2 Estimators for the Variance Parameter

We divide the analysis into three portions: Derivations for IR, derivations for BM, and then a comparison.

4.2.1 Independent Replications

Under Model (10), we can write $\bar{Y}_{i,m} = \bar{T}_{i,m} + \bar{a}_m$, for $i = 1, 2, \ldots, b$, and $\bar{Y}_n = \bar{T}_n + \bar{a}_m$. This immediately implies that

$$\widehat{V}_R = \frac{m}{b-1} \sum_{i=1}^{b} (\bar{Y}_{i,m} - \bar{Y}_n)^2 = \frac{m}{b-1} \sum_{i=1}^{b} (\bar{T}_{i,m} - \bar{T}_n)^2.$$

Since $\bar{T}_{1,m}, \bar{T}_{2,m}, \ldots, \bar{T}_{b,m}$ are i.i.d. r.v.'s with mean μ and variance σ_m^2/m, nothing changes from the steady-state case studied in Section 2.2 — Equation (1) still gives

$$\mathsf{E}[\widehat{V}_R] \;=\; m\mathsf{Var}(\bar{Y}_{i,m}) \;=\; m\mathsf{Var}(\bar{T}_{i,m}) \;=\; \sigma_m^2, \qquad (11)$$

and Equation (4) still gives $\mathsf{Var}(\widehat{V}_R)$ — so both $\mathsf{E}[\widehat{V}_R]$ and $\mathsf{Var}(\widehat{V}_R)$ are unaffected by the additive transient function $\{a_\ell\}$. These "lucky" results make sense here because Model (10) is simply the sum of a stationarity process and a deterministic transient process, the latter of which cancels out in the calculation of $\widehat{V}_R$.

4.2.2 Batch Means

Unlike the case for independent replications, the BM estimator $\widehat{V}_B$ for σ^2 is affected by the transient in Model (10). After some algebra,

$$\widehat{V}_B \;=\; \frac{m}{b-1}\Biggl\{\sum_{i=1}^{b}(\bar{S}_{i,m}-\bar{S}_n)^2 + 2\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)\bar{S}_{i,m} + \sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2\Biggr\},$$

This leads to the following results, analogous to Theorems 1 and 2.

Theorem 4 If Model (10) holds and all necessary moments exist, then

$$\mathsf{E}[\widehat{V}_B] \;=\; \sigma^2+\frac{\gamma(b+1)}{n}+o\left(\frac{1}{m}\right)+\frac{m}{b-1}\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2. \qquad (12)$$

Theorem 5 Suppose the process $\{X_\ell\}$ satisfies Model (10) with bounded transient constants $\{a_\ell\}$ such that $a_n = o(1)$. Further suppose that the process $\{S_\ell\}$ is stationary with $\mathsf{E}[S_1^{12}] < \infty$ and ϕ-mixing with $\phi_k = O(k^{-9})$. Then

$$b\,\mathsf{Var}(\widehat{V}_B) \;=\; 2\sigma^4+\frac{4n\sigma_m^2}{(b-1)^2}\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2+o(1). \qquad (13)$$

Remark 2 Suppose, as would be the case for large batch size m, that the batch means are approximately independent normal r.v.'s. Then under Model (10), a result adapted from Equation (5) of Goldsman, Schruben, and Swain (1994) shows that

$$\widehat{V}_B \;\approx\; \frac{\sigma_m^2}{b-1}\,\chi^2\!\left(b-1,\; \frac{m}{\sigma_m^2}\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2\right),$$

where $\chi^2(\nu,\delta)$ denotes the noncentral χ^2 distribution with ν degrees of freedom and noncentrality parameter δ. Using well-known moment properties of the noncentral χ^2, we can retrieve Equations (12) and (13).

4.2.3 Comparison

For Model (10), we can compare the expression for $\mathsf{E}[\widehat{V}_B]$ given by Equation (12) with that for $\mathsf{E}[\widehat{V}_R]$, which is still given by Equation (1). Assuming that $\sigma_n^2 \doteq \sigma_m^2$, we have

$$\mathsf{E}[\widehat{V}_B] \;\doteq\; \mathsf{E}[\widehat{V}_R]+\frac{m}{b-1}\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2,$$

and thus $\widehat{V}_B$ has the potential for a great deal of positive (conservative) bias as an estimator of σ^2. As explained in Remark 1 and Section 3.2, $\widehat{V}_R$ and $\widehat{V}_B$ are often biased for σ^2 from below, at least for processes with positive autocorrelation; so $\mathsf{E}[\widehat{V}_B]$'s extra term is not necessarily deleterious for purposes of estimating σ^2.

Similarly, we can compare $\mathsf{Var}(\widehat{V}_B)$ from Equation (13) with $\mathsf{Var}(\widehat{V}_R)$, which is still given by Equation (4). Under Model (10),

$$\mathsf{Var}(\widehat{V}_B) \;\doteq\; \mathsf{Var}(\widehat{V}_R)+\frac{4m\sigma_m^2}{(b-1)^2}\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{a}_n)^2.$$

Again the additional noncentrality term appears, indicating that $\mathsf{Var}(\widehat{V}_B)$ tends to be higher than $\mathsf{Var}(\widehat{V}_R)$ for our simple Model (10). Combining the above bias and variance results shows that it is likely, but not always certain, that $\mathsf{MSE}(\widehat{V}_B;\sigma^2) > \mathsf{MSE}(\widehat{V}_R;\sigma^2)$.

Remark 3 It may very well be the case that, under a different transient than Model (10), BM will outperform IR in terms of the bias of the respective variance estimators. See Alexopoulos and Goldsman (2003) for such examples.

4.3 Confidence Intervals for the Mean

As in Section 3.3, we encounter difficulties with respect to the analysis on CI's for μ. So *for purposes of simplifying the rough-cut analysis*, we shall assume that the replicate and batch means are normally distributed with the appropriate parameters.

4.3.1 Independent Replications

Taking advantage of the liberal assumptions of this subsection, we can derive the probability of coverage for the IR

CI for μ under Model (10). First of all, Alexopoulos and Goldsman (2003) show that the pivot

$$T_R^\star \equiv \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\widehat{V}_R^{1/2}} \sim t\left(b-1\,,\ \frac{\sqrt{bm}\,\bar{a}_m}{\sigma_m}\right), \qquad (14)$$

where $t(\nu, \delta)$ is the noncentral t distribution with ν degrees of freedom and noncentrality parameter δ (cf. Evans, Hastings, and Peacock 2000, Chapter 39). Thus, the probability that the IR CI covers μ is

$$\begin{aligned}&\Pr\left(\mu \in \bar{Y}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_R/n}\right)\\ &\quad = F_{T_R^\star}(t_{\alpha/2,b-1}) - F_{T_R^\star}(-t_{\alpha/2,b-1}), \qquad (15)\end{aligned}$$

where $F_{T_R^\star}(\cdot)$ is the c.d.f. of $T_R^\star$.

4.3.2 Batch Means

In order to make a rough-cut analysis on the BM CI for μ under Model (10), we will also assume that the batch means $\bar{X}_{i,m}$, $i = 1, 2, \ldots, b$, are approximately independent — probably reasonable for sufficiently large batch size m. Now, Alexopoulos and Goldsman (2003) show that the pivot

$$T_B^\star \equiv \frac{\sqrt{n}(\bar{X}_n - \mu)}{\widehat{V}_B^{1/2}} \approx \frac{\sigma_n}{\sigma_m}\, t\left(b-1,\ \frac{\bar{a}_n\sqrt{n}}{\sigma_n},\ \delta_B\right), \quad (16)$$

where $t(\nu, \delta_1, \delta_2)$ is the doubly noncentral t distribution with ν degrees of freedom and noncentrality parameters δ_1 and δ_2 (cf. Krishman 1968), and

$$\delta_B \equiv \frac{m}{\sigma_m^2}\sum_{i=1}^{b}(\bar{a}_{i,m} - \bar{a}_n)^2.$$

Thus, under Model (10), the probability that the BM CI covers μ is

$$\begin{aligned}&Pr\left(\mu \in \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n}\right)\\ &\quad \doteq G\left(\frac{\sigma_m}{\sigma_n}t_{\alpha/2,b-1}\right) - G\left(-\frac{\sigma_m}{\sigma_n}t_{\alpha/2,b-1}\right),\end{aligned}$$

where $G(\cdot)$ is the c.d.f. of the doubly noncentral t random variable in Equation (16).

4.3.3 Comparison

One cannot make completely sweeping conclusions regarding the comparative performance of the IR and BM CI's for the steady-state mean. Nevertheless, some interesting findings are possible.

- For Model (10) with a fixed number of batches b, the effect of the batch size m on the IR CI's coverage depends on the form of the sequence $\{a_j\}$. In particular, the t distribution noncentrality parameter from Equation (14), $\sqrt{bm}\,\bar{a}_m/\sigma_m$, will converge to zero if the underlying a_ℓ's approach zero sufficiently quickly, e.g., $a_\ell = o(1/\sqrt{\ell})$. In that case, the CI's coverage will approach the nominal value $1-\alpha$ as m increases. If the a_ℓ's approach zero more slowly, coverage degradation may result; in fact, it may very well be the case that $\sqrt{bm}\,\bar{a}_m/\sigma_m$ approaches some non-zero constant, whence the coverage may never converge to the nominal value! See Section 5.
- For Model (10) with fixed m, the IR CI's noncentrality parameter $\sqrt{bm}\,\bar{a}_m/\sigma_m$ increases in the number of replications b. One would expect a resulting adverse effect on the coverage of the IR CI; this is borne out in some additional examples given in Alexopoulos and Goldsman (2003).
- The BM method's first noncentrality parameter $\sqrt{n}\,\bar{a}_n/\sigma_n$ from Equation (16) behaves qualitatively similarly to the corresponding IR parameter $\sqrt{n}\,\bar{a}_m/\sigma_m$; but since the a_ℓ's converge to zero, the BM's noncentrality parameter will likely be closer to zero than that of IR — a potentially huge advantage for BM. Not as much can be said about the behavior of the BM method's second noncentrality parameter δ_B in Equation (16), nor its effects on CI coverage.

4.4 Transient Recap

With respect to point estimation of the steady-state mean μ, the results from Section 4.1 indicate that an initial transient is more likely to be a problem for IR than for BM — particularly in terms of bias when the underlying process follows Model (10). Section 4.2 shows that the comparison between the IR and BM estimators for the steady-state parameter σ^2 is somewhat inconclusive. On the other hand, Section 4.3 hints strongly that, when it comes to CI estimation for μ, there may be more problems on the horizon for IR than for BM.

5 EXAMPLE

This section illustrates our findings with an example involving the first-order autoregressive [AR(1)] process.

We start off with the stationary AR(1) process

$$S_\ell = \rho S_{\ell-1} + \xi_\ell, \quad \ell \geq 1,$$

where the ξ_ℓ's are i.i.d. Nor$(0, 1-\rho^2)$ r.v.'s with $\rho \in (-1, 1)$ and $S_0 \sim$ Nor$(0, 1)$. For this process, $R_i = \rho^{|i|}$, and some easy calculations (see, e.g., Sargent, Kang, and Goldsman 1992) give

$$\sigma_m^2 \;=\; \sigma^2 + \frac{\gamma(1-\rho^m)}{m}$$

with

$$\sigma^2 \;=\; \frac{1+\rho}{1-\rho} \quad \text{and} \quad \gamma \;=\; \frac{-2\rho}{(1-\rho)^2}.$$

We now turn to Model (10), i.e., $X_\ell = S_\ell + a_\ell$, and study the performance of the resulting IR and BM estimators for σ^2. First of all, the discussion in Section 4.2.1 — in particular, Equation (11) — implies that $\mathsf{E}[\widehat{V}_R] = \sigma_m^2$; similarly, Equations (5) and (2) show that

$$(b-1)\,\mathsf{Var}(\widehat{V}_R) \;=\; 2\sigma_m^4 \;=\; 2\sigma^4 + O(1/m).$$

Further, Alexopoulos and Goldsman (2003) derive the following (see also Carlstein 1986, p. 1176).

$$\mathsf{E}[\widehat{V}_B] \;=\; \sigma^2 + \frac{\gamma(b+1)}{n} + O\left(\frac{\rho^m}{m}\right) + \frac{m}{b-1}\sum_{i=1}^{b}(\bar{a}_{i,m} - \bar{a}_n)^2$$

and

$$(b-1)\,\mathsf{Var}(\widehat{V}_B) \;=\; 2\sigma^4 + \frac{4(b+1)\gamma\sigma^2}{n} + O\left(\frac{1}{m^2}\right) + \Psi(\mathbf{a}),$$

where $\Psi(\mathbf{a})$ is a messy function of the a_ℓ's.

As we continue to study the nonstationary process $\{X_\ell\}$, we set $\rho = 0.9$ and consider the initialization functions $a_\ell = 1/\ell^p$, for $p = 1$ and $1/2$. Table 1 contains experimental results comparing the IR and BM methods; the comparison is based on the achieved sample coverage ($\widehat{\mathsf{CVG}}$) of the 95% CI's for the steady-state mean $\mu = 0$ and the estimated expected value ($\widehat{\mathsf{E}}[\widehat{V}]$) of the estimators for the variance parameter $\sigma^2 = 19$. All estimators are based on 10000 independent experiments, each with $b = 20$ independent replications or batches, and various values of m.

We first examine the case $a_\ell = 1/\ell$, a sequence of initialization bias constants that converges to zero relatively quickly. In this case, both the IR and BM CI's for μ appear to achieve the nominal coverage as m increases, with BM succeeding a bit more quickly than IR. The same good behavior holds true for the respective estimators of σ^2. These coverage results are not surprising in light of the fact that all of the IR and BM noncentrality parameters in Sections 4.3.1 and 4.3.2 die to zero as m becomes large; nor are the variance estimation results surprising in light of Equation (11) and Theorem 4.

We have a particularly interesting story for the case $a_\ell = 1/\sqrt{\ell}$, a sequence of bias constants that converges to zero relatively slowly. For $b = 20$, the BM method nearly (but not quite) achieves the nominal coverage, while the coverage of the analogous IR CI's is poor *regardless of the replication size m*. Although the IR coverage results are disappointing, the IR variance estimator nevertheless achieves $\widehat{\mathsf{E}}[\widehat{V}_R] \doteq 19$.

Table 1: Experimental Results for the "Biased" AR(1) Process with $\rho = 0.9$, $b = 20$, and Transient Functions $\{a_\ell\}$ (Standard Errors of All $\widehat{\mathsf{CVG}}$'s Are ≤ 0.003)

	m	$a_\ell = 1/\ell$		$a_\ell = 1/\ell^{1/2}$	
		$\widehat{\mathsf{CVG}}$	$\widehat{\mathsf{E}}[\widehat{V}]$	$\widehat{\mathsf{CVG}}$	$\widehat{\mathsf{E}}[\widehat{V}]$
IR	100	0.918	17.23	0.521	17.23
	500	0.941	18.56	0.514	18.56
	1000	0.947	18.68	0.511	18.68
	2500	0.945	18.77	0.509	18.77
BM	100	0.940	17.14	0.914	17.28
	500	0.949	18.53	0.925	18.67
	1000	0.953	18.67	0.926	18.81
	2500	0.950	18.76	0.927	18.91

This seemingly bizarre behavior of the IR coverage when $p = 1/2$ has an explanation. Simply put, for $a_\ell = 1/\sqrt{\ell}$, the noncentrality parameter of the t distribution in Equation (14) does not converge to zero as the batch size m becomes large; and although not illustrated here, it turns out that the bad effects become more-pronounced as the number of replications b increases. So in this $p = 1/2$ case, the coverage probability in Equation (15) cannot be nominal even if the replication length m becomes large! The same phenomenon also occurs with respect to batch means, but is much less of a problem since there is essentially one long replication consisting of b batches. See Alexopoulos and Goldsman (2003) for all of the surprising details.

Of course, we could attempt to ameliorate these coverage problems by truncating an initial portion of each replication (or the single batch means run), but this also has to be done with extreme care (cf. Fishman 2001, Section 3.4).

6 CONCLUSIONS

In this paper, we presented a comparison between the IR and BM methods. The comparison was based on several new results as well as on illustrative examples. We focused on nonstationary models with an additive transient, and under the assumption that both methods use the same pair (b, m).

When the process under study is in steady state (or the transient portion has been removed successfully), the IR and BM estimators for μ and σ^2 are practically equivalent with regard to their MSE's as the replication/batch size m becomes large. However, in the steady-state case, the IR method wins with regard to the coverage of the CI for μ;

indeed, if the replicate means are normal, the IR coverage is exactly nominal. On the other hand, the typical presence of an initial transient turns the tide in favor of the BM method. This assessment — for the transient case — is based on the slower convergence of the respective IR-based estimators for μ. See Alexopoulos and Goldsman (2003) for additional examples.

REFERENCES

Alexopoulos, C., and D. Goldsman. 2003. To batch or not to batch? Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Alexopoulos, C., and A. F. Seila. 1998. Output data analysis. In *Handbook of Simulation*, ed. J. Banks, Chapter 7. New York: John Wiley and Sons.

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

Carlstein, E. 1986. The use of subseries for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* 14:1171–1179.

Chien, C.-H., D. Goldsman, and B. Melamed. 1997. Large-sample results for batch means. *Management Science* 43:1288–1295.

Evans, M., N. Hastings, and B. Peacock. 2000. *Statistical Distributions*, 3rd ed. New York: John Wiley and Sons.

Fishman, G.S. 2001. *Discrete-Event Simulation: Modeling, Programming, and Analysis*. New York: Springer-Verlag.

Glynn, P. W., and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.

Glynn, P. W., and W. Whitt. 1991. Estimating the asymptotic variance with batch means. *Operations Research Letters* 10:431–435.

Goldsman, D., and M. S. Meketon. 1986. A comparison of several variance estimators. Technical Report #J-85-12, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Goldsman, D., L. W. Schruben, and J. J. Swain. 1994. Tests for transient means in simulated time series. *Naval Research Logistics* 41:171–187.

Kang, K., and D. Goldsman. 1990. The correlation between mean and variance estimators in computer simulation. *IIE Transactions* 22:15–23.

Krishman, M. 1968. Series representation of the doubly non-central *t* distribution. *Journal of the American Statistical Association* 63(323):1004–1012.

Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. New York: McGraw-Hill.

Sargent, R. G., K. Kang, and D. Goldsman. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Operations Research* 40:898–913.

Song, W.-M., and B. W. Schmeiser. 1995. Optimal mean-squared-error batch sizes. *Management Science* 41:110–123.

Titus, B. D. 1985. Modified confidence intervals for the mean of an autoregressive process. Technical Report No. 34, Department of Operations Research, Stanford University, Stanford, CA.

Whitt, W. 1991. The efficiency of one long run versus independent replications in steady-state simulation. *Management Science* 37:645–666.

AUTHOR BIOGRAPHIES

CHRISTOS ALEXOPOULOS is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. in Operations Research from the University of North Carolina at Chapel Hill. His research interests are in the areas of applied probability, statistics, and optimization of stochastic systems. His recent work involves problems related to the optimal design of telecommunications and supply chain systems. Dr. Alexopoulos is a member of INFORMS and the INFORMS College on Simulation. He served as a Co-Editor for the *Proceedings of the 1995 Winter Simulation Conference*. His email address is <christos@isye.gatech.edu> and his web page is <http://www.isye.gatech.edu/~christos>.

DAVID GOLDSMAN is a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. in Operations Research and Industrial Engineering from Cornell University. His research interests include simulation output analysis and ranking and selection. Dave serves as the Simulation Department Editor of *IIE Transactions*. He is an active participant in the Winter Simulation Conference, having been Program Chair in 1995, and having served on the WSC Board of Directors since 2002. His email address is <sman@isye.gatech.edu> and his web page is <http://www.isye.gatech.edu/~sman>.