Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

VARIANCE ESTIMATION USING REPLICATED BATCH MEANS

Sigrún Andradóttir
Nilay Tanık Argon

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

ABSTRACT

We present a new method for obtaining confidence intervals in steady-state simulation. In our *replicated batch means* method, we do a small number of independent replications to estimate the steady-state mean of the underlying stochastic process. In order to obtain a variance estimator, we further group the observations from these replications into non-overlapping batches. We show that for large sample sizes, the new variance estimator is less biased than the batch means variance estimator, the variances of the two variance estimators are approximately equal, and the new steady-state mean estimator has a smaller variance than the batch means estimator when there is positive serial correlation between the observations. For small sample sizes, we compare our replicated batch means method with the (standard) batch means and multiple replications methods empirically, and show that the best overall coverage of confidence intervals is obtained by the replicated batch means method with a small number of replications.

1 INTRODUCTION

The problem of constructing a confidence interval on the steady-state mean μ of a stationary process $\{X_i\}_{i\geq 1}$ often arises in simulations. This requires a point estimator for μ and also an estimator for the variance of that point estimator. The usual point estimator for μ is the sample mean, $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, where n is the total number of observations. The more challenging part of constructing a confidence interval is estimating the variance of the sample mean, or equivalently, $\sigma_n^2 = nVar(\bar{X}_n)$. If the number of observations n is large, then the experimenter can instead estimate the *variance parameter*, $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$.

There is a significant amount of research on the problem of obtaining a "good" estimator for σ^2 in the discrete-event simulation literature (see, e.g., Alexopoulos and Seila, 1998). Two of the most popular methods for estimating σ^2 are the method of batch means (BM) and the method of multiple independent replications (MR). With the BM method, the observations obtained from one long run are grouped into non-overlapping batches to estimate σ^2. With the MR method, on the other hand, σ^2 is estimated based on the replication means obtained from the independent replications (which usually start in the same initial conditions). We provide a more detailed background on these methods in Section 2.

The trade-off between making a single long run and many independent replications is studied by many authors in the simulation literature (see, e.g., Kelton and Law, 1984, Whitt, 1991, and the references therein). Most of these studies provide supporting evidence in favor of making one long run (which is the case for the BM method). This is in part due to the fact that the MR method has more data that is contaminated by initialization bias than the BM method (due to using initial conditions that do not represent the long-run behavior of the underlying stochastic process). The most familiar method for dealing with the initialization bias is to truncate some of the initial observations from each replication. This implies that the MR method wastes more data than the BM method.

However, the MR method also has some benefits. One advantage of the MR method is that the replication means are independent. For the BM method, on the other hand, the batch means are usually positively correlated. Another advantage of using multiple replications is that there is a chance to start the individual replications in different initial states and observe various different sample paths of the underlying stochastic process. This is especially useful for stochastic processes such as nearly decomposable Markov chains. In such situations, if the single run of the BM method is not long enough, the experimenter may not detect any abnormalities in the process. With several independent replications, on the other hand, there is a chance that the special structure of the Markov chain might be caught.

In this study, our objective is to introduce a new confidence interval estimation method that combines the advantages of the BM and MR methods. For this purpose, we

suggest collecting a total of n observations in $k > 1$ independent replications. We assume that k is small so that we cannot simply use the replication means to obtain a variance estimator. Therefore, after collecting the data in k independent replications, we further partition the observations in each replication (possibly after accounting for initialization bias by truncating some of the initial observations) into $b > 1$ batches, each consisting of m observations. Hence, the total number of batches will be kb. We refer to this method as the *replicated batch means* (RBM) method.

Since our RBM method involves doing more than one replication, it has fewer correlated batches than the BM method and has the potential to observe different sample paths of the underlying stochastic process that start in different initial conditions. Moreover, since the RBM method uses a smaller number of replications than the MR method, its steady-state mean estimator is likely to be less biased than the steady-state mean estimator obtained by the MR method. Also, with the RBM method, less data will be deleted (due to initialization bias) than with the MR method.

This paper is organized as follows. In Section 2, we provide background material on the BM and MR methods. In Section 3, we introduce the RBM method and present our asymptotic results for the estimators for μ and σ^2. We compare these results to the corresponding results for the BM method in Section 4. In Section 5, we present numerical results for the RBM, BM, and MR methods applied to a single example. Finally, we give our concluding remarks in Section 6.

2 BACKGROUND

2.1 Batch Means (BM)

In Sections 2, 3, and 4 of this paper, we assume that the underlying stochastic process $\{X_i\}_{i\geq 1}$ is stationary. The original BM method makes a single long run and then partitions the observations $X_1, \ldots, X_n$ from this run into $kb > 1$ non-overlapping batches, each consisting of m observations (we assume that $n = kbm$). Then,

$$\bar{X}_j = \frac{1}{m}\sum_{i=1}^{m} X_{(j-1)m+i}$$

is the j^{th} batch mean, where $j = 1, \ldots, kb$, $\bar{X}_{BM} = \sum_{i=1}^{n} X_i/n$ is the point estimator for the steady-state mean μ, and

$$\widehat{V}_{BM} = \frac{m}{kb-1}\sum_{j=1}^{kb}\left(\bar{X}_j - \bar{X}_{BM}\right)^2$$

is the BM estimator for $\sigma_n^2 = nVar(\bar{X}_{BM})$. When the number of batches kb is fixed and m is large, we can assume that the batch means are approximately i.i.d. (independent and identically distributed) normal random variables (see, e.g., Glynn and Iglehart 1990). Therefore, an approximate $100(1-\alpha)\%$ confidence interval for μ is given by

$$\mu \in \bar{X}_{BM} \pm t_{kb-1,\alpha/2}\sqrt{\widehat{V}_{BM}/n},$$

where $t_{d,\alpha/2}$ is the $1-\alpha/2$ quantile of the t distribution with d degrees of freedom.

2.2 Multiple Replications (MR)

For the multiple independent replications method, kb independent replications of the underlying stochastic process, each of length m observations, are performed to estimate μ and σ^2. We define $X_i^{(r)}$ to be the i^{th} observation from the r^{th} replication for $i = 1, \ldots, m$ and $r = 1, \ldots, kb$. We also let

$$\bar{Y}_r = \frac{1}{m}\sum_{i=1}^{m} X_i^{(r)}$$

be the r^{th} replication mean, where $r = 1, \ldots, kb$. This yields the point estimator $\bar{X}_{MR} = \sum_{r=1}^{kb}\bar{Y}_r/kb$ for μ. A point estimator for σ^2 can be obtained by the sample variance,

$$\widehat{V}_{MR} = \frac{m}{kb-1}\sum_{r=1}^{kb}\left(\bar{Y}_r - \bar{X}_{MR}\right)^2.$$

Finally, note that when m is large, we can assume that $\bar{Y}_1, \ldots, \bar{Y}_{kb}$ are approximately i.i.d. normal random variables. Therefore, an approximate $100(1-\alpha)\%$ confidence interval for μ is given by

$$\mu \in \bar{X}_{MR} \pm t_{kb-1,\alpha/2}\sqrt{\widehat{V}_{MR}/n}.$$

3 REPLICATED BATCH MEANS (RBM)

As we mentioned earlier, the RBM method involves running a small number k of independent replications of the underlying stochastic process and grouping the observations in each replication into b non-overlapping batches of size m. Let

$$\bar{X}_j^{(r)} = \frac{1}{m}\sum_{i=1}^{m} X_{(j-1)m+i}^{(r)}$$

be the j^{th} batch mean of the r^{th} replication, where $r = 1, \ldots, k$ and $j = 1, \ldots, b$, $\bar{X}_{RBM} = \sum_{r=1}^{k}\sum_{j=1}^{bm} X_j^{(r)}/n$ be the point estimator for μ, and

$$\widehat{V}_{RBM} = \frac{m}{kb-1}\sum_{r=1}^{k}\sum_{j=1}^{b}\left(\bar{X}_j^{(r)} - \bar{X}_{RBM}\right)^2$$

be the RBM estimator for σ^2. For stationary processes, we can assume that the batch means obtained from each replication are approximately i.i.d. normal random variables (the batch means from different replications are always independent). With this assumption, we obtain the following approximate $100(1-\alpha)\%$ confidence interval for μ,

$$\mu \in \bar{X}_{RBM} \pm t_{kb-1,\alpha/2}\sqrt{\widehat{V}_{RBM}/n}$$

(for a rigorous proof and explicit conditions, we refer the reader to Andradóttir and Argon 2001).

Before we state our results on the asymptotic expectation and variance of $\bar{X}_{RBM}$ and $\widehat{V}_{RBM}$ for a stationary process $\{X_i\}_{i\geq 1}$, we define the covariance function $R_j = Cov(X_1, X_{1+j})$, $j = 1, 2, \ldots$, and the relevant quantity $\gamma = -2\sum_{j=1}^{\infty} jR_j$. We also use the notation $g(n) = O(f(n))$ to express that $|g(n)/f(n)| \leq C$ for some constant C and all $n \geq 1$, and $g(n) = o(f(n))$ to indicate that $g(n)/f(n) \to 0$ as $n \to \infty$.

We first study the properties of the point estimator, $\bar{X}_{RBM}$. Since the stochastic process under consideration is assumed to be stationary, $\bar{X}_{RBM}$ is an unbiased estimator for the steady-state mean (i.e., $E[\bar{X}_{RBM}] = \mu$). Thus, the mean squared error of the estimator, $MSE(\bar{X}_{RBM})$, equals $Var(\bar{X}_{RBM})$. The next result gives an expression for the variance of $\bar{X}_{RBM}$.

Theorem 1. *Under mild moment and mixing conditions (see Andradóttir and Argon 2001), we have*

$$Var\left(\bar{X}_{RBM}\right) = \frac{\sigma^2}{kbm} + \frac{\gamma}{kb^2m^2} + o\left(\frac{1}{m^2}\right).$$

We next present our results on the expected value, variance, and mean squared error of the RBM variance estimator, $\widehat{V}_{RBM}$, for large m.

Theorem 2. *Under the same conditions as in Theorem 1, we have*

$$E\left[\widehat{V}_{RBM}\right] = \sigma^2 + \frac{\left(kb^2-1\right)\gamma}{(kb-1)\,bm} + o\left(\frac{1}{m}\right).$$

Theorem 2 shows that, as $m \to \infty$, $\widehat{V}_{RBM}$ is an asymptotically unbiased estimator for σ^2, regardless of the choice of k and b.

Theorem 3. *Under slightly more restrictive moment and mixing conditions than the ones needed to prove Theorems 1 and 2 (see Andradóttir and Argon 2001), we have*

$$\begin{aligned} kbVar\left(\widehat{V}_{RBM}\right) &= \frac{2kb\,(kb+1)\,\sigma^4}{(kb-1)^2} + O\left(\frac{1}{m^{1/4}}\right) \\ &\quad + O\left(\frac{1}{kb}\right) \\ &= 2\sigma^4 + o(1), \end{aligned} \tag{1}$$

as $m \to \infty$ and $kb \to \infty$.

From Theorems 2 and 3, we have

$$MSE\left(\widehat{V}_{RBM}\right) = \frac{\gamma^2}{m^2} + \frac{2\sigma^4}{kb} + o\left(\frac{1}{m^2}\right) + o\left(\frac{1}{kb}\right). \tag{2}$$

Hence, for large m and kb, the estimator $\widehat{V}_{RBM}$ is consistent in mean square.

4 BM vs. RBM

In this section, we compare our results obtained in Section 3 with the corresponding results for the BM method. First, we note that by the stationarity of the underlying stochastic process, both $\bar{X}_{BM}$ and $\bar{X}_{RBM}$ are unbiased estimators for the steady-state mean (i.e., $E[\bar{X}_{RBM}] = E[\bar{X}_{BM}] = \mu$). Therefore, comparing the variances of $\bar{X}_{BM}$ and $\bar{X}_{RBM}$ is equivalent to comparing their mean squared errors. From Song and Schmeiser (1995) and Theorem 1 above, we get

$$Var\left(\bar{X}_{BM}\right) - Var\left(\bar{X}_{RBM}\right) = \frac{(1-k)\gamma}{k^2b^2m^2} + o\left(\frac{1}{m^2}\right).$$

Thus, if $\gamma < 0$ (e.g., if $R_j \geq 0$ for all $j \in \{1, 2, \ldots\}$, corresponding to a stochastic process with positive serial correlation), then the RBM method offers a point estimator with a smaller variance than the point estimator obtained by the BM method for large m. This is a notable result, since positive correlation occurs frequently in simulations (for example, the waiting time processes in various queueing systems exhibit positive correlation).

We now turn our attention to the confidence intervals for μ obtained by the BM and RBM methods. The *half-length* of a generic $100(1-\alpha)\%$ confidence interval is defined by $H = t_{d,1-\alpha/2}(\widehat{V}/n)^{1/2}$, where d is the number of degrees of freedom, $\widehat{V}$ is the variance estimator, and n is the total number of observations. Among confidence interval estimation methods achieving coverage of approximately $1-\alpha/2$, the one with the smallest $E[H]$ is preferred, and then that with the smallest $Var(H)$. For both the BM and RBM methods, we have $d = kb - 1$ and $n = kbm$. Hence, the only term that is different in the half-lengths

of the confidence intervals obtained by the BM and RBM methods is the variance estimator. Therefore, a comparison of the variance estimators obtained by the two methods for large m will also yield insights about the half-length estimators in the resulting confidence intervals.

We conclude this section by comparing the BM and RBM methods in terms of expected value, variance, and mean squared error of the variance estimator for large m. By Theorem 1 of Chien, Goldsman, and Melamed (1997) and Theorem 2 above, we get

$$\left|Bias\left(\widehat{V}_{BM}\right)\right| - \left|Bias\left(\widehat{V}_{RBM}\right)\right| = \frac{|\gamma|\,(k-1)}{kbm(kb-1)} + o\left(\frac{1}{m}\right),$$

which is positive for large m. Hence, the RBM method offers an asymptotically less biased variance estimator than the BM method. Consequently, the RBM method can be expected to provide an asymptotically less biased half-length estimator than the BM method. This is important, since a less biased half-length helps in achieving the desired coverage of the corresponding confidence interval. Finally, under the conditions of Theorem 3, Chien, Goldsman, and Melamed (1997) show that for large m and kb, $Var(\widehat{V}_{BM})$ and $MSE(\widehat{V}_{BM})$ are equal to the right-hand sides of equations (1) and (2), respectively. Thus, we can conclude that for a stationary process and large m and kb, the BM and RBM methods will yield variance estimators with approximately the same variance and mean squared error.

5 NUMERICAL RESULTS

In this section, we study the small-sample behavior of the RBM, BM, and MR methods. For this purpose, we consider the first-order autoregressive process with exponential marginals, EAR(1). In particular, for all $i \geq 1$, we have

$$X_i = \begin{cases} \phi X_{i-1} & \text{with probability} \quad \phi, \\ \phi X_{i-1} + \epsilon_i & \text{with probability} \quad 1-\phi, \end{cases}$$

where $0 \leq \phi < 1$ and $\{\epsilon_i\}_{i\geq 1}$ are i.i.d. exponential random variables with rate one (see Lewis, 1980). Our primary criterion for comparing the RBM, BM, and MR methods is the coverage $P\{\mu \in \bar{X} \pm H\}$ of the resulting confidence intervals, where $\bar{X}$ is the point estimator for μ and H is the half-length of the confidence interval (we say that a good coverage is achieved when $P\{\mu \in \bar{X} \pm H\}$ is very close to $1-\alpha$, the desired confidence level).

We have simulated the EAR(1) process for $\phi \in \{0.1, 0.5, 0.9\}$ and examined the coverages and the expected half-lengths of the confidence intervals under the BM, RBM, and MR methods. For each method, we let $m = 4$ and $kb = 32$. Consequently, the BM method has 32 batches, the RBM method with $k \in \{2, 4, 8, 16\}$ replications has $32/k$ batches per replication, and the MR method has 32 replications with one batch per replication (we have not truncated any initial observations).

In Tables 1, 2, and 3, we provide the estimated coverage values when X_0 is distributed uniformly between 0 and $a \in \{0.00, 0.25, 0.50, \ldots, 5.00\}$. Noting that the stationary distribution is exponential with mean one, it is clear that we have a relatively good choice of the initial distribution when a is close to two, and that when a is close to zero or five, then we have a poor choice of the initial distribution. All the numerical results that are presented in this section are based on 10,000 independent macro replications. For each macro replication, we have used common random numbers to generate observations from the EAR(1) process for the three methods. In the last two rows of Tables 1, 2, and 3, we also provide the averages and the standard deviations of the coverages over all initial distributions for each method.

Table 1: Coverages of Confidence Intervals for the EAR(1) Process with $\phi = 0.1$ and $1-\alpha = 0.90$

a	BM	RBM				MR
		$k=2$	$k=4$	$k=8$	$k=16$	
0.00	0.8816	0.8817	0.8803	0.8825	0.8783	0.8652
0.25	0.8816	0.8817	0.8804	0.8822	0.8797	0.8760
0.50	0.8817	0.8815	0.8811	0.8823	0.8816	0.8806
0.75	0.8816	0.8814	0.8813	0.8825	0.8817	0.8834
1.00	0.8817	0.8816	0.8814	0.8824	0.8822	0.8867
1.25	0.8818	0.8817	0.8818	0.8824	0.8843	0.8918
1.50	0.8817	0.8818	0.8821	0.8827	0.8847	0.8957
1.75	0.8818	0.8820	0.8821	0.8837	0.8856	0.8997
2.00	0.8819	0.8821	0.8827	0.8846	0.8854	0.9010
2.25	0.8821	0.8821	0.8831	0.8854	0.8862	0.9034
2.50	0.8819	0.8822	0.8836	0.8854	0.8875	0.9055
2.75	0.8819	0.8822	0.8838	0.8856	0.8883	0.9060
3.00	0.8820	0.8825	0.8845	0.8862	0.8890	0.9045
3.25	0.8821	0.8828	0.8846	0.8865	0.8896	0.9038
3.50	0.8820	0.8829	0.8848	0.8872	0.8908	0.9033
3.75	0.8822	0.8830	0.8854	0.8873	0.8913	0.9013
4.00	0.8824	0.8831	0.8858	0.8874	0.8919	0.8991
4.25	0.8824	0.8837	0.8859	0.8879	0.8906	0.8957
4.50	0.8827	0.8838	0.8859	0.8883	0.8898	0.8929
4.75	0.8830	0.8841	0.8864	0.8881	0.8885	0.8909
5.00	0.8829	0.8842	0.8865	0.8878	0.8878	0.8869
avg.	0.8820	0.8825	0.8835	0.8852	0.8864	0.8940
std.	0.0004	0.0009	0.0021	0.0023	0.0040	0.0109

First, we note that for all values of ϕ that we have considered in this study, the BM, RBM, and MR methods all yield under-coverage (except for the MR method with a good choice of the initial distribution and $\phi = 0.1$), which is partially due to the small sample size and the positive correlation inherent in the EAR(1) process. For $\phi = 0.1$, the coverages of the BM, RBM, and MR methods for each value of a are very similar, as can be seen from Table 1. This is an expected result, since with a small value of ϕ, the EAR(1) process behaves almost like an i.i.d. stochastic process.

In Tables 2 and 3, for each choice of a, we have highlighted the coverages of the method with the best coverage.

Table 2: Coverages of Confidence Intervals for the EAR(1) Process with $\phi = 0.5$ and $1 - \alpha = 0.90$

a	BM	RBM				MR
		$k = 2$	$k = 4$	$k = 8$	$k = 16$	
0.00	**0.8170**	0.8142	0.8059	0.7786	0.6763	0.3969
0.25	**0.8172**	0.8165	0.8093	0.7893	0.7078	0.4679
0.50	0.8172	**0.8176**	0.8120	0.7989	0.7351	0.5479
0.75	0.8177	**0.8191**	0.8152	0.8087	0.7612	0.6303
1.00	0.8179	**0.8200**	**0.8182**	0.8148	0.7889	0.7049
1.25	0.8193	**0.8214**	**0.8214**	**0.8244**	0.8114	0.7730
1.50	0.8198	0.8234	0.8241	0.8286	**0.8312**	0.8288
1.75	0.8208	0.8222	0.8265	0.8350	0.8455	**0.8664**
2.00	0.8212	0.8231	0.8281	0.8413	0.8556	**0.8926**
2.25	0.8213	0.8236	0.8292	0.8430	0.8627	**0.8987**
2.50	0.8221	0.8244	0.8307	0.8438	0.8628	**0.8846**
2.75	0.8228	0.8262	0.8335	0.8446	**0.8572**	0.8496
3.00	0.8229	**0.8279**	**0.8360**	**0.8441**	**0.8476**	0.7959
3.25	0.8231	**0.8288**	**0.8355**	**0.8419**	**0.8344**	0.7228
3.50	0.8234	**0.8303**	**0.8357**	**0.8389**	0.8126	0.6362
3.75	0.8243	**0.8317**	**0.8361**	**0.8343**	0.7913	0.5388
4.00	0.8257	**0.8317**	**0.8370**	**0.8309**	0.7657	0.4439
4.25	0.8263	**0.8311**	**0.8372**	**0.8280**	0.7363	0.3498
4.50	0.8267	**0.8323**	**0.8375**	0.8207	0.7008	0.2683
4.75	0.8276	**0.8328**	**0.8373**	0.8128	0.6645	0.1958
5.00	0.8278	**0.8334**	**0.8376**	0.8045	0.6230	0.1402
avg.	0.8220	**0.8253**	**0.8278**	**0.8241**	0.7796	0.6111
std.	0.0035	0.0059	0.0103	0.0193	0.0731	0.2421

Table 3: Coverages of Confidence Intervals for the EAR(1) Process with $\phi = 0.9$ and $1 - \alpha = 0.90$

a	BM	RBM				MR
		$k = 2$	$k = 4$	$k = 8$	$k = 16$	
0.00	**0.4736**	0.4644	0.3969	0.2074	0.0351	0.0009
0.25	**0.4747**	0.4689	0.4188	0.2542	0.0589	0.0033
0.50	**0.4763**	0.4745	0.4436	0.3083	0.1003	0.0093
0.75	0.4777	**0.4798**	0.4691	0.3690	0.1658	0.0296
1.00	0.4788	**0.4844**	**0.4934**	0.4321	0.2626	0.0811
1.25	0.4793	**0.4918**	**0.5116**	**0.4997**	0.3856	0.2004
1.50	0.4783	**0.4973**	**0.5265**	**0.5623**	**0.5354**	0.4216
1.75	0.4802	0.4999	0.5404	0.6165	0.6788	**0.7057**
2.00	0.4814	0.5034	0.5542	0.6485	0.7715	**0.8847**
2.25	0.4835	0.5066	0.5597	0.6563	0.7666	**0.8367**
2.50	0.4858	0.5085	0.5590	**0.6383**	**0.6666**	0.6222
2.75	0.4863	**0.5103**	**0.5548**	**0.5862**	**0.5180**	0.3583
3.00	0.4899	**0.5139**	**0.5480**	**0.5205**	0.3606	0.1686
3.25	0.4898	**0.5116**	**0.5297**	0.4489	0.2348	0.0668
3.50	0.4943	**0.5109**	**0.5132**	0.3774	0.1468	0.0262
3.75	0.4959	**0.5090**	0.4934	0.3090	0.0895	0.0099
4.00	0.4976	**0.5098**	0.4684	0.2511	0.0535	0.0030
4.25	0.5013	**0.5057**	0.4389	0.2014	0.0324	0.0008
4.50	0.5030	**0.5061**	0.4137	0.1579	0.0198	0.0002
4.75	0.5022	**0.5037**	0.3846	0.1243	0.0121	0.0001
5.00	**0.5051**	0.5023	0.3559	0.0994	0.0065	0.0000
avg.	0.4874	**0.4982**	0.4845	0.3937	0.2810	0.2109
std.	0.0103	0.0150	0.0638	0.1847	0.2705	0.3018

For $\phi = 0.5$ and $\phi = 0.9$, it can be observed that the RBM method with a small number of replications k, is overall the best in terms of the average coverage (recall that our initial intention was to apply the RBM method with a small number of replications). Moreover, the overall performance of the BM method is better than that of the MR method.

When we observe each row of Tables 2 and 3, we can see that our RBM method performs better than both the BM and MR methods for reasonably good choices of the initial distribution. This suggests that the RBM method is more than an *average* of the BM and MR methods. More specifically, the RBM method shows the best performance for $\phi = 0.5$ when $a \in \{0.50, \ldots, 1.50, 2.75, \ldots, 5.00\}$ and it is the best for $\phi = 0.9$ when $a \in \{0.75, \ldots, 1.50, 2.50, \ldots, 4.75\}$. Moreover, the RBM method never performs worse than both the BM and MR methods. As one might expect, the MR method performs better than the other two methods when the initial distribution is quite good (i.e., when $a \in \{1.75, 2.00, 2.25, 2.50\}$ for $\phi = 0.5$ and $a \in \{1.75, 2.00, 2.25\}$ for $\phi = 0.9$). On the other hand, in this example the BM method provides the best coverage for very poor choices of initial distribution (i.e., when $a \in \{0.00, 0.25\}$ for $\phi = 0.5$ and when $a \in \{0.00, 0.25, 0.50, 5.00\}$ for $\phi = 0.9$).

From the last row of Tables 2 and 3, we can see that the BM method and the RBM method with a small number of replications provide smaller standard deviations of the coverage than the other methods. This suggests that the BM method and the RBM method with a small number of replications are not very sensitive to the choice of the initial distribution, whereas the MR method and the RBM method with a large number of replications show performance that depends heavily on the choice of the initial distribution. For example, depending on the choice of the initial distribution, the MR method yields both the best and the worst results among all the methods.

We have also observed the performance of the BM, RBM, and MR methods starting at various initial distributions in terms of the resulting half-lengths. Our conclusion is that there is not much of a difference among the BM, RBM, and MR methods in this respect. To illustrate this point, we provide the average half-lengths of the confidence intervals over all the initial distributions considered in Tables 1, 2, and 3 in Table 4.

Table 4: Average Half-lengths of Confidence Intervals for the EAR(1) Process with $1 - \alpha = 0.90$

ϕ	BM	RBM				MR
		$k = 2$	$k = 4$	$k = 8$	$k = 16$	
0.1	0.1591	0.1591	0.1591	0.1591	0.1592	0.1595
0.5	0.2107	0.2109	0.2112	0.2115	0.2114	0.2085
0.9	0.2486	0.2517	0.2543	0.2534	0.2468	0.2368

6 CONCLUSIONS

We have presented a new confidence interval estimation method for steady-state simulation. In our *replicated batch means* (RBM) method, we do a small number of independent replications to estimate the steady-state mean and then group the observations that are collected from these replica-

tions into non-overlapping batches to estimate the variance parameter.

We have shown that for a stationary process and large batch sizes, the RBM method provides a less biased variance estimator than the batch means (BM) method, regardless of the choice of the number of batches and replications. Moreover, we have shown that for a stationary process, the variances of the two variance estimators become approximately equal as both the number of batches and batch sizes grow. Finally, we have shown that when there is positive serial correlation in the stationary process of interest, then the point estimator obtained by the RBM method has a smaller variance than the point estimator obtained by the BM method for large batch sizes. This is important, since most real-life queueing systems produce positively correlated data.

We have also compared the performance of the BM, RBM, and MR methods for small sample sizes, using a non-stationary stochastic process. The numerical results suggest that our RBM method with a small number of replications provides better coverages than both the BM and MR methods for reasonably good choices of the initial distribution. The MR method performs the best only for good choices of the initial distribution, and, in our example, the BM method appears to have the best coverage for very poor choices of the initial distribution, especially when the observations have high positive autocorrelation.

REFERENCES

Alexopoulos, C., and A. F. Seila. 1998. Output Data Analysis. Chapter 7 in the *Handbook of Simulation–Principles, Methodology, Advances, Applications, and Practice*, ed. J. Banks, 225–272. New York: John Wiley and Sons, Inc.

Andradóttir, S., and N. T. Argon. 2001. Replicated batch means for steady-state simulation. Working paper.

Chien, C. H., D. Goldsman, and B. Melamed. 1997. Large-sample results for batch means. *Management Science* 43 (9): 1288–1295.

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

Kelton, W. D., and A. M. Law. 1984. An analytical evaluation of alternative strategies in steady-state simulation. *Operations Research* 32 (1): 169–184.

Lewis, P. A. W. 1980. Simple models for positive-valued and discrete-valued time series with ARMA correlation structure. In *Multivariate Analysis–V: Proceedings of the Fifth International Symposium on Multivariate Analysis*, ed. P. R. Krishnaiah, 151–166. New York: North-Holland.

Song, W.-M. T., and B. W. Schmeiser. 1995. Optimal mean-squared-error batch sizes. *Management Science* 41 (1): 110–123.

Whitt, W. 1991. The efficiency of one long run versus independent replications in steady-state simulation. *Management Science* 37 (6): 645–666.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants DMI–9523111 and DMI–0000135.

AUTHOR BIOGRAPHIES

SIGRÚN ANDRADÓTTIR is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Previously, she was an Assistant Professor and later an Associate Professor in the Departments of Industrial Engineering, Mathematics, and Computer Sciences at the University of Wisconsin – Madison. She received a B.S. in Mathematics from the University of Iceland in 1986, an M.S. in Statistics from Stanford University in 1989, and a Ph.D. in Operations Research from Stanford University in 1990. Her research interests include simulation, applied probability, and stochastic optimization. She is a member of INFORMS and served as Editor of the *Proceedings of the 1997 Winter Simulation Conference*.

NİLAY TANIK ARGON is a Ph.D. student in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. She received her B.S. and M.S. degrees in Industrial Engineering from the Middle East Technical University, Ankara, Turkey, in 1996 and 1998, respectively. In August 2000, she received an M.S. degree in Operations Research from the Georgia Institute of Technology. Her primary research interests are the modeling and analysis of stochastic systems and simulation output analysis. She is a member of INFORMS. Her e-mail address is `nilay@isye.gatech.edu`.