Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

IMPLEMENTING THE BATCH MEANS METHOD IN SIMULATION EXPERIMENTS

Christos Alexopoulos

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

Andrew F. Seila

Terry College of Business
University of Georgia
Athens, GA 30602–6255, U.S.A.

ABSTRACT

This paper reviews and evaluates strategies for implementing the batch means method for estimating the mean of a stationary simulation output process.

1 INTRODUCTION

Suppose $\{X_i, i \geq 1\}$ is a discrete-time stochastic process. The method of batch means is frequently used to estimate the steady-state mean μ of $\{X_i\}$ or the $\mathsf{Var}(\overline{X}_n)$ (for finite n) and owes its popularity to its simplicity and effectiveness. The original references on the method are Conway (1963), Fishman (1978a,b), and Law and Carson (1979).

The classical approach divides the output $X_1, \ldots, X_n$ of a long simulation run into a number of contiguous *batches* and uses the sample means of these batches (or *batch means*) to produce point and interval estimators.

To motivate the method, suppose temporarily that the process $\{X_i\}$ is weakly stationary, that is, $\mathsf{E}(X_i) = \mu$, $\mathsf{Var}(X_i) = \sigma^2$, and the $\mathsf{Cov}(X_i, X_j)$ depends only on the *lag* $|j - i|$. Also assume that $\lim_{n\to\infty} n\mathsf{Var}(\overline{X}_n) < \infty$. Then split the data into k batches, each consisting of b observations. (Assume $n = kb$.) The ith batch consists of the observations

$$X_{(i-1)b+1}, X_{(i-1)b+2}, \ldots, X_{ib}$$

for $i = 1, 2, \ldots, k$ and the ith batch mean is given by

$$Y_i(b) = \frac{1}{b}\sum_{j=1}^{b} X_{(i-1)b+j}.$$

For fixed m, let $\sigma_m^2 = \mathsf{Var}(\overline{X}_m)$. Since the batch means process $\{Y_i(b), i \geq 1\}$ is also weakly stationary, some algebra yields

$$\begin{aligned}\sigma_n^2 &= \frac{\sigma_b^2}{k} + \frac{1}{k^2}\sum_{i\neq j}\mathsf{Corr}[Y_i(b), Y_j(b)] \\ &= \frac{\sigma_b^2}{k}\left(1 + \frac{n\sigma_n^2 - b\sigma_b^2}{b\sigma_b^2}\right). \qquad (1)\end{aligned}$$

Since $n \geq b$, $(n\sigma_n^2 - b\sigma_b^2)/(n\sigma_b^2) \to 0$ as first $n \to \infty$ and then $b \to \infty$. As a result, σ_b^2/k approximates σ_n^2 with error that diminishes as b and n approach infinity. Equivalently, the correlation among the batch means diminishes as b and n approach infinity.

To use the last limiting property, one forms the grand batch mean

$$\overline{Y}_k = \overline{X}_n = \frac{1}{k}\sum_{i=1}^{k} Y_i(b),$$

estimates σ_b^2 by

$$\hat{V}_B(n,k) = \frac{1}{k-1}\sum_{i=1}^{k}(Y_i(b) - \overline{Y}_k)^2,$$

and computes the following approximate $1-\alpha$ confidence interval for μ:

$$\overline{Y}_k \pm t_{k-1,1-\alpha/2}\sqrt{\hat{V}_B(n,k)/k}. \qquad (2)$$

The main problem with the application of the batch means method in practice is the choice of the batch size b. If b is too small, the means $Y_i(b)$ can be highly correlated and the resulting confidence interval will frequently have coverage below the user-specified nominal coverage $1-\alpha$. Alternatively, a large batch size will likely result in very few batches and potential problems with the application of the central limit theorem to obtain (2). An extensive study of batch size effects for fixed sample size was conducted by Schmeiser (1982).

Remark 1 For fixed sample size, a plot of the batch means is a very useful tool for checking the effects of initial conditions, non-normality of batch means, and existence of correlation between batch means. For example, consider the M/M/1 queueing system with interarrival rate $\tau = 0.09$ and service rate $\omega = 0.1$. The limiting mean customer delay is $\mu = \tau/[\omega(\omega-\tau)] = 90$. A sample of 100,000 customer

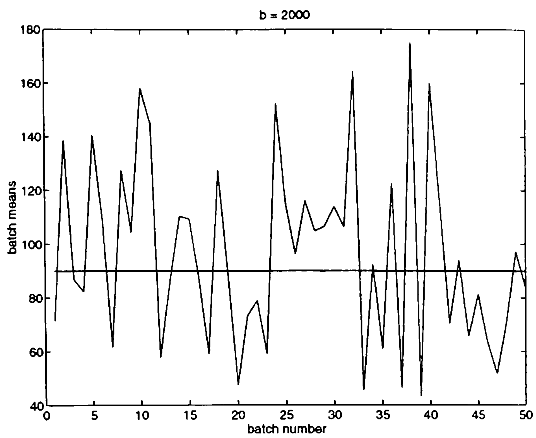

Figure 1: Batch means for delay times in an M/M/1 queue

delays, D_i, was generated by means of Lindley's recursion $D_{i+1} = \max\{D_i+S_i-A_{i+1},0\}$, $i \geq 1$, starting with an empty system ($D_1 = 0$), where A_i is the ith interarrival time and S_i is the service time of the ith customer. For moderate-to-heavy traffic intensity $\nu = \tau/\omega$, the autocorrelation function for the process $\{D_i\}$ has a very long tail (see Blomqvist 1967). This property makes the M/M/1 system a good model for evaluating simulation methodologies.

Figure 1 shows the plot of the batch means $Y_1(2000), \ldots, Y_{50}(2000)$ for batch size $b = 2000$. The first batch mean is small but not the smallest, relaxing one's worries about the effect of the initial transient period. This also hints that $l = 2000$ is a reasonable truncation index for removing transient observations. Had the first batch mean been smaller than the other batch means, one can assess the effect of the initial conditions by removing the first batch and comparing the new grand batch mean with the old. Although the plot does not indicate the presence of serious autocorrelation among the batch means, the asymmetric dispersion of the batch means about the actual mean should make the experimenter concerned about the coverage of the confidence interval (2).

Example 1 shows how an asymptotically optimal batch size can be obtained in special cases.

Example 1 Consider the stationary AR(1) process

$$X_i = \mu + \rho(X_{i-1} - \mu) + Z_i, \quad i \geq 1,$$

where $|\rho| < 1$, $X_0 \sim N(\mu, 1)$, and the Z_i's are i.i.d. $N(0, 1-\rho^2)$. Carlstein (1986) showed that

$$\mathsf{Bias}(\hat{V}_B(n,k)) = -\frac{2\rho}{(1-\rho)^3(1+\rho)}\frac{1}{b} + o\left(\frac{1}{b}\right) \quad (3)$$

and

$$\mathsf{Var}(\hat{V}_B(n,k)) = \frac{2}{(1-\rho)^4}\frac{b}{n} + o\left(\frac{b}{n}\right),$$

where $o(h)$ is a function such that $\lim_{h\to 0} o(h)/h = 0$. Then the batch size that minimizes the asymptotic (as $n \to \infty$ and $k \to \infty$) mean squared error $\mathsf{MSE}(\hat{V}_B(n,k)) = \mathsf{Bias}^2(\hat{V}_B(n,k)) + \mathsf{Var}(\hat{V}_B(n,k))$ is

$$b_0 = \left(\frac{2|\rho|}{1-\rho^2}\right)^{2/3} n^{1/3}. \quad (4)$$

Clearly, the optimal batch size increases with the absolute value of the correlation ρ between successive observations. Unfortunately, such an analysis cannot be performed for the majority of output processes. Furthermore, asymptotically optimal batch sizes may differ considerably from optimal batch sizes for finite sample sizes (as Song and Schmeiser (1995) observed for a congested M/M/1 queue).

2 CONSISTENT ESTIMATION BATCH MEANS METHODS

Consistent estimation batch means methods assume the existence of a parameter σ^2 (the time-average variance of the process $\{X_i\}$), such that a central limit theorem holds

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \sigma N(0,1) \quad \text{as } n \to \infty \quad (5)$$

and aim at constructing a consistent estimator for σ^2 and an asymptotically valid confidence interval for μ. [Notice that the X_i's in (5) need not be i.i.d.] Consistent estimation methods are often preferable to methods that "cancel" σ^2 (see Glynn and Iglehart 1990) because: (a) The expectation and variance of the halfwidth of the confidence interval resulting from (5) is asymptotically smaller for consistent estimation methods; and (b) Under reasonable assumptions $n\mathsf{Var}(\overline{X}_n) \to \sigma^2$ as $n \to \infty$.

Chien, Goldsman, and Melamed (1996) considered stationary processes and, under quite general moment and sample path conditions, showed that as both $b, k \to \infty$, $\mathsf{MSE}[b\hat{V}_B(n,k)] \to 0$. Notice that mean squared error consistency differs from consistency.

The limiting result (5) is implied under the following two assumptions, where $\{W(t), t \geq 0\}$ is the standard Brownian motion process (see Resnick 1994, Chapter 6).

Assumption of Weak Approximation (AWA). There exist finite constants μ and $\sigma > 0$ such that

$$\frac{n(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} W(n) \quad \text{as } n \to \infty.$$

Assumption of Strong Approximation (ASA). There exist finite constants μ, $\sigma > 0$, $\lambda \in (0, 1/2]$, and a finite random variable C such that, with probability one,

$$|n(\overline{X}_n - \mu) - \sigma W(n)| \leq Cn^{1/2-\lambda} \quad \text{as } n \to \infty.$$

Both AWA and ASA state that the process $\{n(\overline{X}_n - \mu)/\sigma\}$ is close to a standard Brownian motion. However the stronger ASA addresses the convergence rate of (5).

The ASA is not restrictive as it holds under relatively weak assumptions for a variety of stochastic processes including Markov chains, regenerative processes and certain queueing systems (see Damerdji 1994 for details). The constant λ is closer to 1/2 for processes having little autocorrelation while it is closer to zero for processes with high autocorrelation. In the former case the "distance" between the processes $\{n(\overline{X}_n - \mu)/\sigma\}$ and $\{W(n)\}$ "does not grow" as n increases.

2.1 Batching Rules

Fishman and Yarberry (1994) and Fishman (1996, Chapter 6) presented a thorough discussion of batching rules. Both references contain detailed instructions for obtaining FORTRAN implementations for various platforms via anonymous ftp.

Equation (1) suggests that fixing the number of batches and letting the batch size grow as $n \to \infty$ ensures that $\sigma_b^2/k \to \sigma_n^2$. This motivates the following rule.

The Fixed Number of Batches (FNB) Rule. Fix the number of batches at k. For sample size n, use batch size $b_n = \lfloor n/k \rfloor$.

The FNB rule along with AWA lead to the following result.

Theorem 1 (Glynn and Iglehart 1990) *If $\{X_i\}$ satisfies AWA, then as $n \to \infty$, $\overline{X}_n \xrightarrow{p} \mu$ and (5) holds. Furthermore, if k is constant and $\{b_n, n \geq 1\}$ is a sequence of batch sizes such that $b_n \to \infty$ as $n \to \infty$, then*

$$\frac{\overline{X}_n - \mu}{\sqrt{\hat{V}_B(n,k)/k}} \xrightarrow{d} t_{k-1} \quad \text{as } n \to \infty.$$

The primary implication of Theorem 1 is that

$$\overline{Y}_k \pm t_{k-1,1-\alpha/2}\sqrt{\hat{V}_B(n,k)/k} \qquad (6)$$

is an asymptotically valid confidence interval for μ. Unfortunately, the FNB rule has two major limitations: (a) $b_n \hat{V}_B(n,k)$ is not a consistent estimator of σ^2. Therefore the confidence interval (6) tends to be wider than the interval a consistent estimation method would produce; (b) Statistical fluctuations in the halfwidth of the confidence interval (6) do not diminish relative to statistical fluctuation in the sample mean (see Fishman 1996, pp. 544–545).

The following theorem proposes batching assumptions which along with ASA yield a strongly consistent estimator for σ^2.

Theorem 2 (Damerdji 1994) *If $\{X_i\}$ satisfies ASA, then $\overline{X}_n \xrightarrow{a.s.} \mu$ as $n \to \infty$. Furthermore suppose that $\{(b_n, k_n), n \geq 1\}$ is a batching sequence satisfying*

1. *$b_n \to \infty$ and $k_n \to \infty$ monotonically as $n \to \infty$*
2. *$b_n^{-1} n^{1-2\lambda} \ln n \to 0$ as $n \to \infty$*
3. *there exists a finite positive integer a such that*

$$\sum_{n=1}^{\infty} (b_n/n)^a < \infty.$$

Then

$$b_n \hat{V}_B(n, k_n) \xrightarrow{a.s.} \sigma^2 \qquad (7)$$

and

$$Z_{k_n} = \frac{\overline{X}_n - \mu}{\sqrt{\hat{V}_B(n,k_n)/k_n}} \xrightarrow{d} N(0,1). \qquad (8)$$

The last display implies that

$$\overline{X}_n \pm t_{k_n-1,1-\alpha/2}\sqrt{\hat{V}_B(n,k_n)/k_n}$$

is an asymptotically valid $1-\alpha$ confidence interval for μ.

Theorem 2 motivates the consideration of batch sizes of the form $b_n = \lfloor n^\theta \rfloor$, $0 < \theta < 1$. In this case one can show that the conditions (1)–(3) are met if $\theta \in (1-2\lambda, 1)$. In particular, the assignment $\theta = 1/2$ and the SQRT rule below are valid if $1/4 < \lambda < 1/2$. Notice that the last inequality is violated by processes having high autocorrelation ($\lambda \approx 0$).

The Square Root (SQRT) Rule. For sample size n, use batch size $b_n = \lfloor \sqrt{n} \rfloor$ and number of batches $k_n = \lfloor \sqrt{n} \rfloor$.

Under some additional moment conditions, Chien (1989) showed that the convergence of Z_{k_n} to the $N(0,1)$ distribution is fastest if both b_n and k_n grow proportionally to $\sqrt{n}$. Unfortunately, in practice the SQRT rule tends to seriously underestimate the $\mathsf{Var}(\overline{X}_n)$ for fixed n.

Example 2 (The M/M/1 queue) Consider an M/M/1 queueing system with interarrival rate $\tau = 0.9$ and service rate $\omega = 1$, and assume that the system starts empty. Table 1 contains performance statistics for 0.95 confidence intervals for the steady-state mean customer delay $\mu = 9$. The confidence intervals resulted from 500 independent replications. The FNB rule used 16 batches and batch sizes 2^m, $m \geq 0$. The SQRT rule started with batch size $b_1 = 1$ and number of batches $k_1 = 8$, and computed confidence intervals with batch sizes

$$b_l = 2^{(l-1)/2} \times \begin{cases} b_1 & \text{if } l \text{ is odd} \\ 3/(2\sqrt{2}) & \text{otherwise} \end{cases}$$

and numbers of batches

$$k_l = 2^{(l-1)/2} \times \begin{cases} k_1 & \text{if } l \text{ is odd} \\ 11/\sqrt{2} & \text{otherwise.} \end{cases}$$

Columns 2 and 4 contain the estimated coverage probabilities of the confidence intervals produced by the FNB rule and the SQRT rule, respectively. Columns 3 and 5 display the respective average interval halfwidths. Specifically, for sample size $n \approx 2^{17} =$ 131,072, roughly 94 percent of the confidence intervals resulting from the FNB rule contained μ whereas only 78 percent of the confidence intervals resulting from the SQRT rule contained μ. However, the latter intervals were 43 percent narrower. Experiments by Fishman and Yarberry showed that the disparity in coverage between the two rules grows with increasing traffic intensity.

Table 1: Performance statistics for the FNB and SQRT rules on 0.95 confidence intervals for the mean customer delay in an M/M/1 queue with utilization $\nu = 0.9$

	FNB Rule		SQRT Rule	
$\log_2 n$	Coverage	Average Halfwidth	Coverage	Average Halfwidth
10	0.544	3.244	0.326	1.694
11	0.640	3.506	0.366	1.665
12	0.746	3.304	0.414	1.437
13	0.798	2.963	0.466	1.271
14	0.838	2.435	0.498	1.063
15	0.880	1.901	0.604	0.904
16	0.912	1.437	0.664	0.738
17	0.944	1.053	0.778	0.599
18	0.934	0.756	0.810	0.471
19	0.950	0.541	0.854	0.369
20	0.940	0.385	0.858	0.283

With the contrasts between the FNB and SQRT rules in mind, Fishman and Yarberry proposed two strategies that dynamically shift between the two rules. Both strategies perform "interim reviews" and compute confidence intervals at times $n_l \approx n_1 2^{l-1}, l = 1, \ldots, L+1$.

The LBATCH Strategy. At time n_l, if an hypothesis test detects autocorrelation between the batch means, the batching for the next review is determined by the FNB rule. If the test fails to detect correlation, all future reviews omit the test and employ the SQRT rule.

The ABATCH Strategy. If at time n_l the hypothesis test detects correlation between the batch means, the next review employs the FNB rule. If the test fails to detect correlation, the next review employs the SQRT rule.

Both strategies LBATCH and ABATCH yield random sequences of batch sizes. Under relatively mild assumptions, these sequences imply convergence results analogous to (7) and (8) (see Fishman and Yarberry 1994 and Fishman 1996).

Test for Correlation

We will briefly review a test for the hypothesis H_0: the batch means $Y_1(b), \ldots, Y_k(b)$ are uncorrelated. A commonly used test is due to von Neumann (1941) and is very effective when the number of batches k is small.

Assume that the process $\{X_i\}$ is weakly stationary and let

$$\rho_l(b) = \mathrm{Corr}[Y_i(b), Y_{i+l}(b)], \quad l = 0, 1, \ldots$$

be the autocorrelation function of the batch means process. The von Neumann test statistic for H_0 is

$$R_k = \sqrt{\frac{k^2-1}{k-2}} \left[\hat{\rho}_1(b) + \frac{(Y_1(b) - \overline{X}_n)^2 + (Y_k(b) - \overline{X}_n)^2}{2\sum_{i=1}^{k}(Y_i(b) - \overline{X}_n)^2} \right],$$

where

$$\hat{\rho}_1(b) = \frac{\sum_{i=1}^{k-1}(Y_i(b) - \overline{X}_n)(Y_{i+1}(b) - \overline{X}_n)}{\sum_{i=1}^{k}(Y_i(b) - \overline{X}_n)^2}$$

is an estimator for the lag-1 autocorrelation $\rho_1(b)$.

Under H_0, $R_k \approx N(0,1)$ for large b (the batch means become approximately normal) or large k (by the central limit theorem). If $\{X_i\}$ has a monotone decreasing autocorrelation function (e.g., the delay process for an M/M/1 queueing system), one rejects H_0 at level β if $R_k > z_{1-\beta}$. Alternatively, if $\{X_i\}$ has an autocorrelation function with damped harmonic behavior around the zero axis (e.g., an AR(1) process with $\rho < 0$), the rejection of H_0 when $R_k > z_{1-\beta}$ can lead to erroneous conclusions. In this case, repeated testing under the ABATCH strategy reduces this possibility.

The p-value $= 1 - \Phi(R_k)$ of the test is the largest value of the type I error $\beta = P(\text{reject } H_0 \mid H_0 \text{ is true})$ given the observed value of R_k. Hence, a p-value close to zero implies low credibility for H_0. The plot of the p-values versus the batch size is a useful graphical device.

2.2 Implementing the LBATCH and ABATCH Strategies

To understand the role of the hypothesis test in the LBATCH and ABATCH algorithms, define the random variables

$$\overline{R}_l = \text{fraction of rejected tests for } H_0 \text{ on reviews } 1, \ldots, l.$$

A sufficient condition for strong consistency (equation (7)) and asymptotic normality (equation (8)) is $\beta_0 > 1 - 4\lambda$ (or $\lambda > (1 - \beta_0)/4$), where $\beta_0 = \lim_{l\to\infty} \overline{R}_l$ is the long-run fraction of rejections. In practice, β_0 differs from but is expected to be close to the type I error β. Clearly, $\lambda > 1/4$ guarantees (7) and (8) regardless of β_0. However, β_0 plays a small role when $\lambda \le 1/4$. Specifically, for β_0 equal to 0.05 or 0.10, the lower bound $(1 - \beta_0)/4$ becomes 0.2375 or 0.2225, respectively, a small reduction from 1/4.

On review l, both strategies induces batch size

$$b_l = 2^{(l-1)(1+\overline{R}_{l-1})/2} \times \begin{cases} b_1 & \text{if } (l-1)(1+\overline{R}_{l-1}) \text{ is even} \\ \tilde{b}_1/\sqrt{2} & \text{otherwise,} \end{cases}$$

where

$$\tilde{b}_1 = \begin{cases} 3/2 & \text{if } b_1 = 1 \\ \lfloor \sqrt{2}\, b_1 + 0.5 \rfloor & \text{if } b_1 > 1, \end{cases}$$

and number of batches

$$k_l = 2^{(l-1)(1-\overline{R}_{l-1})/2} \times \begin{cases} k_1 & \text{if } (l-1)(1-\overline{R}_{l-1}) \text{ is even} \\ \tilde{k}_1/\sqrt{2} & \text{otherwise,} \end{cases}$$

where $\tilde{k}_1 = \lfloor \sqrt{2}\, k_1 + 0.5 \rfloor$.

The resulting sample sizes are

$$n_l = k_l b_l = \begin{cases} 2^{l-1} k_1 b_1 & \text{if } (l-1)(1+\overline{R}_{l-1}) \text{ is even} \\ 2^{l-2} \tilde{k}_1 \tilde{b}_1 & \text{otherwise} \end{cases}$$

and the definitions for $\tilde{b}_1$ and $\tilde{k}_1$ guarantee that if H_0 is never rejected, then both b_l and k_l grow approximately as $\sqrt{2}$ with l (i.e., they follow the SQRT rule).

The final implementation issue is the relative difference between the potential terminal sample sizes

$$\Delta(b_1, k_1) = \frac{|2^L k_1 b_1 - 2^{L-1} \tilde{k}_1 \tilde{b}_1|}{2^L k_1 b_1} = \frac{|2 k_1 b_1 - \tilde{k}_1 \tilde{b}_1|}{2 k_1 b_1}.$$

This quantity is minimized (i.e., the final sample size is deterministic) when $2k_1 b_1 = \tilde{k}_1 \tilde{b}_1$. Although this condition excludes several practical choices for b_1 and k_1, such as $b_1 = 1$ (to test the original sample for independence) and $8 \le k_1 \le 10^5$, $\Delta(b_1, k_1)$ remains small for numerous choices of b_1 and k_1.

2.3 Tests for the Batching Rules

The experiments in Examples 3, 4 and 5 compare the LBATCH and ABATCH strategies by means of three queueing systems with traffic intensity $\nu = 0.9$. Each system starts empty and has a first-come, first-served discipline. Each experiment computed 0.95 confidence intervals for the long-run mean customer delay from 500 independent replications. Both strategies started with $k_1 = 8$ batches of size $b_1 = 1$ and used type I error $\beta = 0.1$ for H_0.

Example 3 (Example 2 continued) The entries of Tables 1 and 2 indicate that the ABATCH strategy comes closer to the FNB rule's superior coverage with shorter confidence intervals.

Table 2: Performance statistics for the LBATCH and ABATCH strategies on 0.95 confidence intervals for the mean customer delay in an M/M/1 queue with utilization $\nu = 0.9$

	LBATCH Strategy		ABATCH Strategy	
$\log_2 n$	Coverage	Average Halfwidth	Coverage	Average Halfwidth
10	0.398	2.085	0.562	3.384
11	0.420	1.992	0.632	3.450
12	0.464	1.693	0.712	3.100
13	0.518	1.477	0.760	2.686
14	0.562	1.227	0.816	2.168
15	0.652	1.029	0.850	1.708
16	0.714	0.834	0.902	1.296
17	0.808	0.663	0.932	0.955
18	0.852	0.513	0.938	0.688
19	0.866	0.395	0.930	0.493
20	0.876	0.298	0.936	0.353

Example 4 (An M/G/1 queue) Consider an M/G/1 queueing system with i.i.d. interarrival times from the exponential distribution with parameter $\tau = 0.9$ and i.i.d. service times S_i from the hyperexponential distribution with density function

$$f(x) = 0.9\left(\frac{1}{0.5}e^{-x/0.5}\right) + 0.1\left(\frac{1}{5.5}e^{-x/5.5}\right), \quad x \ge 0.$$

This distribution applies when customers are classified into two types, 1 and 2, with respective probabilities 0.9 and 0.1, type 1 customers have exponential service times with mean 0.5, and type 2 customers have exponential service times with mean 5.5. The service times have mean $\mathsf{E}(S) = 0.9(0.5) + 0.1(5.5) = 1$, second moment $\mathsf{E}(S^2) = 0.9(0.5^2) + 0.1(5.5^2) = 6.5$, and coefficient of variation $\sqrt{\mathsf{Var}(S)}/\mathsf{E}(S) = 2.345$, which is larger than the unit coefficient of variation of the exponential distribution.

The long-run mean delay time in queue is given by the Pollaczec-Khintchine formula (Heyman and Sobel

1982, pp. 250–252)

$$\mu = \lim_{i\to\infty} \mathsf{E}(D_i) = \frac{\tau \mathsf{E}(S^2)}{2[1-\tau \mathsf{E}(S)]} = 29.25. \quad (9)$$

Notice that the M/M/1 system in Example 2 with the same traffic intensity $\nu = 0.9$ has much smaller long-run mean delay time.

Table 3 displays the results of this experiment. As n increases, the conservative ABATCH strategy produces 0.95 confidence intervals for μ that are roughly 50 to 100 percent wider than the respective confidence intervals produced by the LBATCH strategy but have coverage rates that are acceptably close to 0.95 for substantially smaller sample sizes (as small as 2^{17} = 131,072).

Table 3: Performance statistics for the LBATCH and ABATCH strategies on 0.95 confidence intervals for the mean customer delay in an M/G/1 queue with hyperexponential service times and utilization ν = 0.9

	LBATCH Strategy		ABATCH Strategy	
$\log_2 n$	Coverage	Average Halfwidth	Coverage	Average Halfwidth
10	0.204	5.865	0.356	10.305
11	0.254	5.962	0.436	11.426
12	0.294	5.552	0.566	11.635
13	0.354	5.083	0.652	11.166
14	0.392	4.418	0.746	10.147
15	0.452	3.863	0.794	8.658
16	0.540	3.215	0.856	7.057
17	0.620	2.678	0.898	5.483
18	0.632	2.178	0.896	4.090
19	0.694	1.761	0.900	2.997
20	0.748	1.387	0.924	2.145
21	0.806	1.083	0.926	1.525

Example 5 (An M/D/1 queue) Consider an M/G/1 queueing system with i.i.d. interarrival times from the exponential distribution with parameter $\tau = 0.9$ and fixed unit service times. Then, by (9), the long-run mean delay time in queue is $\mu = 4.5$.

The results of this experiment are contained in Table 4. As in Examples 3 and 4, the performance of the ABATCH strategy makes it an attractive compromise between the "extreme" FNB and SQRT rules.

Example 6 tests the LBATCH and ABATCH methods by means of an AR(1) process.

Example 6 Consider the stationary AR(1) process $X_i = -0.9X_{i-1} + Z_i$ with mean 0 (see Example 1). The autocorrelation function $\rho_j = (-0.9)^j$, $j \geq 0$, of this process oscillates around the zero axis and the time-average process variance is $\sigma^2 = (1-0.9)/(1+0.9) = 0.053$.

Table 4: Performance statistics for the LBATCH and ABATCH strategies on 0.95 confidence intervals for the mean customer delay in an M/D/1 queue with unit service times and utilization $\nu = 0.9$

	LBATCH Strategy		ABATCH Strategy	
$\log_2 n$	Coverage	Average Halfwidth	Coverage	Average Halfwidth
10	0.460	1.062	0.616	1.631
11	0.548	0.962	0.720	1.538
12	0.598	0.842	0.788	1.391
13	0.648	0.686	0.842	1.101
14	0.696	0.556	0.858	0.852
15	0.794	0.445	0.884	0.632
16	0.808	0.351	0.924	0.472
17	0.862	0.271	0.942	0.343

Table 5: Performance statistics for the LBATCH and ABATCH strategies on 0.95 confidence intervals for the mean $\mu = 0$ of the stationary AR(1) process $X_i = -0.9X_{i-1} + Z_i$

	LBATCH Strategy		ABATCH Strategy	
$\log_2 n$	Coverage	Average Halfwidth	Coverage	Average Halfwidth
5	1.000	0.1217	1.000	0.1153
6	0.978	0.0364	0.980	0.0367
7	0.980	0.0244	0.980	0.0246
8	0.982	0.0166	0.980	0.0167
9	0.972	0.0111	0.966	0.0111
10	0.984	0.0076	0.980	0.0075
11	0.976	0.0051	0.978	0.0050
12	0.982	0.0035	0.984	0.0035
13	0.964	0.0024	0.962	0.0023
14	0.960	0.0016	0.962	0.0016

The entries of Table 5 were obtained from 500 independent replications. The type I error for H_0 was $\beta = 0.1$. The 0.95 confidence intervals for μ produced by both methods have roughly equal halfwidths and coverages. In fact, almost all cover age estimates are greater than the nominal coverage 0.95. This behavior is due to the fact that $b\mathsf{Var}(\hat{V}_B(n,k))$ tends to overestimate σ^2 (the coefficient of $1/b$ in equation (3) is $2.624 > 0$).

From equation (4), the batch size that minimizes $\mathsf{MSE}(\hat{V}_B(n,k))$ is $b_0 = 113.71$. 500 independent replications with 144 batches of size 114 (sample size 16416) produced 0.95 confidence intervals with estimated coverage 0.958 and average halfwidth 0.0016 — not a substantial improvement over the statistics in the last row of Table 5 (for sample size roughly equal to $2^{14} = 16384$).

3 OVERLAPPING BATCH MEANS

An interesting variation of the traditional batch means method is the method of *overlapping* batch means (OBM) proposed by Meketon and Schmeiser (1984). For given batch size b, this method uses all $n - b + 1$ overlapping batches to estimate μ and $\mathsf{Var}(\overline{X}_n)$. The first batch consists of observations $X_1, \ldots, X_b$, the second batch consists of $X_2, \ldots, X_{b+1}$, etc. The OBM estimator of μ is

$$\overline{Y}_O = \frac{1}{n-b+1} \sum_{i=1}^{k} Y_i(b)$$

and has sample variance

$$\hat{V}_O = \frac{1}{n-b} \sum_{i=1}^{n-b+1} (Y_i(b) - \overline{Y}_O)^2.$$

The following list contains properties of the estimators $\overline{Y}_O$ and $\hat{V}_O$:

(i) The OBM estimator is a weighted average of non-overlapping batch means estimators.

(ii) Asymptotically (as $n, b \to \infty$ and $b/n \to 0$), the OBM variance estimator $\hat{V}_O$ and the non-overlapping batch means variance estimator $\hat{V}_B \equiv \hat{V}_B(n,k)$ have the same expectation. Furthermore,

$$\frac{\mathsf{Var}(\hat{V}_O)}{\mathsf{Var}(\hat{V}_B)} \to \frac{2}{3}.$$

In words, the asymptotic ratio of the mean squared error of $\mathsf{Var}(\hat{V}_O)$ to the mean squared error of $\mathsf{Var}(\hat{V}_B)$ is equal to 2/3 (Meketon and Schmeiser 1984).

(iii) The behavior of $\mathsf{Var}(\hat{V}_O)$ appears to be less sensitive to the choice of the batch size than the behavior of $\mathsf{Var}(\hat{V}_B)$ (Song and Schmeiser 1995, Table 1).

(iv) If $\{X_i\}$ satisfies ASA and $\{(b_n, k_n), n \geq 1\}$ is a sequence that satisfies the assumptions (A.1)-(A.3) in Theorem 2 and

$$\frac{b_n^2}{n} \to 0 \quad \text{as } n \to \infty,$$

then (Damerdji 1994)

$$b_n \hat{V}_O \stackrel{a.s.}{\to} \sigma^2.$$

Song and Schmeiser (1995) considered weakly stationary processes with $\gamma_m = \sum_{j=-\infty}^{\infty} j^m C_j < \infty$ for $m = 0, 1$ and studied batch means variance estimators with

$$\mathsf{Bias}(\hat{V}) = -c_b \gamma_1 \frac{1}{b} + o\left(\frac{1}{b}\right)$$

and

$$\mathsf{Var}(\hat{V}) = c_v \gamma_0^2 \frac{b}{n} + o\left(\frac{b}{n}\right).$$

The constants c_b and c_v depend on the amount of overlapping between the batches. In particular, the estimator $\hat{V}_B$ has $c_b = 1$ and $c_v = 2$, while $\hat{V}_O$ has $c_b = 1$ and $c_v = 4/3$. Then the asymptotic batch size that minimizes $\mathsf{MSE}(\hat{V}) = \mathsf{Bias}^2(\hat{V}) + \mathsf{Var}(\hat{V})$ is

$$b^* = \left(\frac{2c_b^2\gamma_1^2}{c_v\gamma_0^2}\right)^{1/3} n^{1/3}. \tag{10}$$

Song (1996) developed methods for estimating the ratio $(\gamma_1/\gamma_0)^2$ for a variety of processes, including moving average processes and autoregressive processes. Then one can obtain an estimator for b^* by plugging the ratio estimator into equation (10). Sherman (1995) proposed an alternative method that does not rely on the estimation of $(\gamma_1/\gamma_0)^2$.

Remark 2 Welch (1987) noted that both traditional batch means and overlapping batch means are special cases of spectral estimation and, more importantly, suggested that overlapping batch means yield near-optimal variance reduction when one forms sub-batches within each batch and applies the method to the sub-batches. For example, a batch of size 64 is split into 4 sub-batches and the first (overlapping) batch consists of observations $X_1, \ldots, X_{64}$, the second consists of observations $X_{17}, \ldots, X_{80}$, etc.

REFERENCES

Blomqvist, N. 1967. The covariance function of the M/G/1 queueing system. *Skandinavisk Aktuarietidskrift* 50: 157–174.

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

Chien, C.-H. 1989. Small sample theory for steady state confidence intervals. Technical Report No. 37, Department of Operations Research, Stanford University, Palo Alto, California.

Chien, C., Goldsman, D., and B. Melamed. 1996. Large-sample results for batch means. *Management Science*, to appear.

Conway, R. W. 1963. Some tactical problems in digital simulation. *Management Science* 10:47–61.

Damerdji, H. 1994. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* 19:494–512.

Fishman, G. S. 1978a. Grouping observations in digital simulation. *Management Science* 24:510–521.

Fishman, G. S. 1978b. *Principles of Discrete Event Simulation*. John Wiley, New York, New York.

Fishman, G. S. 1996. *Monte Carlo: Concepts, Algorithms and Applications*. Chapman and Hall, New York, New York.

Fishman, G. S., and L. S. Yarberry. 1994. An implementation of the batch means method. Technical Report UNC/OR/TR/93-1, Department of Operations Research, University of North Carolina, Chapel Hill, North Carolina.

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

Heyman, D. P., and M. J. Sobel. 1982. *Stochastic Models in Operations Research, Volume I*. McGraw-Hill, New York, New York.

Law, A. M., and J. S. Carson. 1979. A sequential procedure for determining the length of a steady-state simulation. *Operations Research* 27:1011–1025.

Meketon, M. S., and B. W. Schmeiser. 1984. Overlapping batch means: Something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, eds. S. Sheppard, U. W. Pooch, and C. D. Pegden, 227–230. IEEE, Piscataway, New Jersey.

Resnick, S. I. 1994. *Adventures in Stochastic Processes*. Birkhaüser, Boston, Massachusetts.

Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.

Sherman, M. 1995. On batch means in the simulation and statistical communities. In *Proceedings of the 1995 Winter Simulation Conference*, eds. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 297–302. IEEE, Piscataway, New Jersey.

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

Song, W.-M. T. 1996. On the estimation of optimal batch sizes in the analysis of simulation output. To appear in *European Journal of Operations Research*.

von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference and the variance. *Annals of Mathematical Statistics* 12:367–395.

Welch, P. D. 1987. On the relationship between batch means, overlapping batch means and spectral estimation, In *Proceedings of the 1987 Winter Simulation Conference*, eds. A. Thesen, H. Grant, and W. D. Kelton, 320–323. IEEE, Piscataway, New Jersey.

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 transportation systems. Dr. Alexopoulos is a member of INFORMS and the INFORMS College of Simulation. He served as a Co-Editor for the *Proceedings of the 1995 Winter Simulation Conference*.

ANDREW SEILA is an Associate Professor in the Terry College of Business at the University of Georgia, Athens, Georgia. He received his B.S. in Physics and Ph.D. in Operations Research and Systems Analysis, both from the University of North Carolina at Chapel Hill. Prior to joining the faculty of the University of Georgia, he was a Member of Technical Staff at the Bell Laboratories in Holmdel, New Jersey. His research interests include all areas of simulation modeling and analysis, especially output analysis. He has been actively involved in the *Winter Simulation Conference* since 1977, and served as a Program Chair for the 1994 conference.