Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

COMPUTATIONAL EXPERIENCE WITH THE BATCH MEANS METHOD

Christos Alexopoulos

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

George S. Fishman

Department of Operations Research
University of North Carolina
Chapel Hill, NC 27599–3180, U.S.A.

Andrew F. Seila

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

ABSTRACT

This article discusses implementation issues for the LBATCH and ABATCH batch means procedures of Fishman and Yarberry (1997). Theses procedures dynamically increase the batch size and the number of contiguous batches based on the outcome of a hypothesis test for independence among the batch means. We show that both procedures require $O(n)$ time and $O(\log_2 n)$ space, where n is the desired sample size. Although like complexities are known for static fixed batch size algorithms, the dynamic setting of the LBATCH and ABATCH rules offers an important additional advantage not present in the static approach. As the analysis evolves with increasing sample path length, it allows a user to assess how well the estimated variance of the sample mean stabilizes. This assessment is essential to gauge the quality of the confidence interval for the sample mean. The LABATCH implementation (described in Fishman 1996 and Fishman and Yarberry 1997) of the LBATCH and ABATCH rules is the only computer package that automatically generates the data for this assessment.

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 $\mathrm{Var}(\overline{X}_n)$ (for finite n) and owes its popularity to its simplicity and effectiveness. Original references on the method are Conway (1963), Fishman (1978a), 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, $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, and the $\mathrm{Cov}(X_i, X_j)$ depends only on the *lag* $|j-i|$. Also assume that $\lim_{n\to\infty} n\mathrm{Var}(\overline{X}_n) < \infty$. 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

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

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

$$\begin{aligned}\sigma_n^2 &= \frac{\sigma_b^2}{k} + \frac{1}{k^2}\sum_{i\neq j}\mathrm{Cov}[\overline{X}_i(b), \overline{X}_j(b)] \\ &= \frac{\sigma_b^2}{k}\left(1 + \frac{n\sigma_n^2 - b\sigma_b^2}{b\sigma_b^2}\right).\psi \qquad (1)\end{aligned}$$

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

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

estimates σ_b^2 by

$$\hat{V}_k(b) = \frac{1}{k-1}\sum_{i=1}^{k}(\overline{X}_i(b) - \overline{X}_n)^2,$$

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

$$\overline{X}_n \pm t_{k-1,1-\alpha/2}\sqrt{\hat{V}_k(b)/k}\,.\psi \qquad (2)$$

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 $\overline{X}_i(b)$ 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).

The method of Fishman (1978) selects the smallest batch size from the set $\{1, 2, 4, \ldots, n/8\}$ that passes the test of independence based on von Neumann's statistic (see Section 2.1). A variant of this method was proposed by Schriber and Andrews (1979). Mechanic and McKay (1966) choose a batch size from the set $\{16b_1, 64b_1, 256b_1, \ldots, n/25\}$ (usually $b_1 = 1$) and select the batch size that passes an alternative test for independence. The procedure of Law and Carson (1979) starts with 400 batches of size 2. Then it considers sample sizes that double every two iterations until an estimate for lag-1 correlation among 400 batch means becomes smaller than 0.4 and larger than the estimated lag-1 correlation among 200 batch means. The procedure stops when the confidence interval (2) computed with 40 batches satisfies a relative width criterion. Schmeiser (1982) reviews the above procedures and concludes that selecting between 10 and 30 batches should suffice for most simulation experiments. The major drawback of these methods is their inability to yield a consistent variance estimator.

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

$$\text{Bias}(\hat{V}_k(b)) = -\frac{2\rho}{(1-\rho)^3(1+\rho)} \times \frac{1}{b} + o\left(\frac{1}{b}\right) \quad (3)$$

and

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

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 $\text{MSE}(\hat{V}_k(b)) = \text{Bias}^2(\hat{V}_k(b)) + \text{Var}(\hat{V}_k(b))$ is

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

Clearly, the optimal batch size increases with the absolute value of the correlation ρ between successive observations.

In practice, the relevance of this model is conjectural. First, the optimal batch size may differ substantially from (4) for a finite sample size (e.g., Song and Schmeiser 1995). Second, the model generally does not apply to the analysis of queueing systems data. Third, it is not evident that this strategy for batch size selection allows the space and time complexities achievable by the LBATCH and ABATCH rules for generating an assessment of the stability of the variance of the sample mean.

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

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

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 $n\text{Var}(\overline{X}_n) \to \sigma_\infty^2$ 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 $b, k \to \infty$, $\mathrm{MSE}(b\hat{V}_k(b)) \to 0$. 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 $\sigma_\infty > 0$ such that

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

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

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

Both AWA and ASA state that the process $\{n(\overline{X}_n - \mu)/\sigma_\infty\}$ 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 $\{n(\overline{X}_n - \mu)/\sigma_\infty\}$ and $\{W(n)\}$ "does not grow" as n increases.

2.1 Batching Rules

Fishman and Yarberry (1997) and Fishman (1996, Chapter 6) presented a thorough discussion of batching rules. Both references contain detailed instructions for obtaining FORTRAN, C, and SIMSCRIPT II.5 implementations for various platforms via anonymous ftp from ftp.or.unc.edu.

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 $n \to \infty$, $\overline{X}_n \xrightarrow{\mathcal{P}} \mu$ 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*

$$\frac{\overline{X}_n - \mu}{\sqrt{\hat{V}_k(b)/k}} \xrightarrow{\mathcal{D}} t_{k-1} \quad \text{as } n \to \infty.$$

The primary implication of Theorem 1 is that (2) is an asymptotically valid confidence interval for μ. Unfortunately, the FNB rule has two major limitations: (a) $b_n\hat{V}_k(b)$ is not a consistent estimator of σ_∞^2. Therefore the confidence interval (2) tends to be wider than the interval a consistent estimation method would produce. (b) Statistical fluctuations in the halfwidth of the confidence interval (2) 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, as $n \to \infty$,

$$b_n \hat{V}_{k_n}(b_n) \xrightarrow{a.s.} \sigma_\infty^2 \tag{6}$$

and

$$Z_{k_n} = \frac{\overline{X}_n - \mu}{\sqrt{\hat{V}_{k_n}(b_n)/k_n}} \xrightarrow{\mathcal{D}} N(0,1).\psi \tag{7}$$

The last display implies that

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

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 $\mathrm{Var}(\overline{X}_n)$ for fixed n.

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

The LBATCH Procedure. 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 Procedure. 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 procedures LBATCH and ABATCH yield random sequences of batch sizes. Under relatively mild assumptions, these sequences imply convergence results analogous to (6) and (7).

Test for Correlation

We will briefly review a test for the hypothesis H_0: the batch means $\overline{X}_1(b), \ldots, \overline{X}_k(b)$ are uncorrelated. A commonly used test is due to von Neumann (1941) and is effective when the number of batches k is as small as 8.

Assume that the process $\{X_i\}$ is weakly stationary. The von Neumann test statistic for H_0 is

$$C_k(b) = \sqrt{\frac{k^2-1}{k-2}} \times \left[1 - \frac{\sum_{i=2}^{k}(\overline{X}_i(b) - \overline{X}_{i-1}(b))^2}{\sum_{i=1}^{k}(\overline{X}_i(b) - \overline{X}_n)^2}\right].$$

Under H_0, $C_k(b) \approx N(0,1)$ 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 $C_k(b) > z_{1-\beta}$. 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 $C_k(b) > z_{1-\beta}$ can lead to erroneous conclusions. In this case, repeated testing under the ABATCH procedure reduces this possibility.

The p–value, $1 - \Phi(C_k(b))$, 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 $C_k(b)$. Equivalently, H_0 is rejected of the p–value is larger than β. Hence, a p–value close to zero implies low credibility for H_0.

2.2 Implementing the LBATCH and ABATCH Procedures

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 (6)) and asymptotic normality (equation (7)) 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 (6) and (7) 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 methods induce 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).

Suppose one decides to perform $L+1$ reviews (iterations). The final implementation issue is the relative difference between the potential terminal sample sizes

$$\Delta(b_1,k_1)=\frac{|2^L k_1 b_1 - 2^{L-1}\tilde{k}_1\tilde{b}_1|}{2^L k_1 b_1}=\frac{|2k_1b_1-\tilde{k}_1\tilde{b}_1|}{2k_1b_1}.$$

This quantity is minimized (i.e., the final sample size is deterministic) when $2k_1b_1=\tilde{k}_1\tilde{b}_1$. 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.

Below we are listing algorithm ABATCH. The implementation of procedure LBATCH is simpler. Once H_0 is accepted in step 15, the steps 17–19 are ignored for the remainder of the execution.

Algorithm ABATCH
Source: Fishman and Yarberry (1997) and Fishman (1996, Chapter 6). Minor notational changes have been made.
Input: Minimal number of batches k_1, minimal batch size b_1, desired sample size $n=2^Lk_1b_1$ (L is a positive integer), and confidence level $1-\alpha$.
Output: Sequences of point estimates and confidence intervals for sample sizes $N\le n$.
Method:

1. $b\leftarrow b_1$ and $k\leftarrow l_1$.
2. If $b_1=1$, $\tilde{b}_1\leftarrow 3/2$;
   otherwise $\tilde{b}_1\leftarrow\lfloor\sqrt{2}\,b_1+0.5\rfloor$.
3. $\tilde{k}_1\leftarrow\lfloor\sqrt{2}\,k_1+0.5\rfloor$.
4. $g\leftarrow\tilde{b}_1/b_1$ and $f\leftarrow\tilde{k}_1/k_1$.
5. $i\leftarrow 0$.
6. $\tilde{n}\leftarrow 2^{L-1}\tilde{k}_1\tilde{b}_1$.
   Until $N=n$ or $N=\tilde{n}$:
7. $N\leftarrow kb$.
8. Randomly generate $X_{i+1},\ldots,X_N$.
   Compute:
9. The batch means $\overline{X}_1(b),\ldots,\overline{X}_k(b)$.
10. $\overline{X}_N$ as a point estimate of μ.
11. The sample variance $\hat{V}_k(b)$ of the batch means.
12. The halfwidth $\delta=t_{k-1,1-\alpha/2}\sqrt{\hat{V}_k(b)/k}$ of the confidence interval (2).
13. Print N, k, b, $\overline{X}_N$, $\overline{X}_N-\delta$, $\overline{X}_N+\delta$, $\hat{V}_k(b)$.
14. $i\leftarrow N$.
15. Test $H_0:\overline{X}_1(b),\ldots,\overline{X}_k(b)$ are uncorrelated. Print the p-value of this test.
16. If H_0 is rejected, $b\leftarrow 2b$. (FNB rule)
    If H_0 is accepted:
17. If $b=1$, $b\leftarrow 2$. (FNB rule)
    Otherwise: (SQRT rule)
18. $b\leftarrow bg$ and $k\leftarrow kf$.
19. If $g=\tilde{b}_1/b_1$, $g\leftarrow 2b_1/\tilde{b}_1$ and $f\leftarrow 2k_1/\tilde{k}_1$;
    otherwise $g\leftarrow\tilde{b}_1/b_1$ and $f\leftarrow\tilde{k}_1/k_1$.

3 COMPUTATIONAL COMPLEXITY

We now discuss the computational complexity of the LBATCH and ABATCH methods. Detailed derivations and a pseudo-code are in Chapter 5 of Yarberry (1993).

On each review, both methods require updated $\overline{X}_N$, $\hat{V}_k(b)$, and $C_k(b)$. Let $S_n=\sum_{i=1}^n X_i$ and $S_0=0$. Then S_i can be computed by the recursion

$$S_i=S_{i-1}+X_i,$$

the batch means can be computed from

$$\overline{X}_i(b)=\frac{1}{b}(S_{ib}-S_{(i-1)b}),$$

and $\overline{X}_N=S_N/N$. This approach eliminates the need to store $X_1,\ldots,X_N$.

Now define

$$\begin{aligned}W_i(b)&=b\overline{X}_i(b)\\ Y_k(b)&=\sum_{i=1}^k W_i(b)^2,\psi\\ Z_k(b)&=\sum_{i=2}^k W_i(b)W_{i-1}(b),\end{aligned}\qquad(8)$$

and notice that $\hat{V}_k(b)$ can be expressed in terms of $Y_k(b)$ and S_N^2:

$$\hat{V}_k(b)=\frac{1}{b^2(k-1)}\left[Y_k(b)-\frac{S_N^2}{k}\right].\psi\qquad(9)$$

Furthermore, one can write von Neumann's statistic as

$$C_k(b)=\frac{-S_N^2/k+W_1(b)^2/2+W_k(b)^2/2+Z_k(b)}{Y_k(b)-S_N^2/k}.$$

Let $W_0(b)=0$, $Y_0(b)=0$, and $Z_0(b)=0$. Since

$$\begin{aligned}W_i(b)&=S_{ib}-S_{(i-1)b},\\ Y_i(b)&=Y_{i-1}(b)+W_i(b)^2,\quad\text{and}\\ Z_i(b)&=Z_{i-1}(b)+W_i(b)W_{i-1}(b),\end{aligned}\qquad(10)$$

the quantities $\hat{V}_k(b)$ and $C_k(b)$ can be computed from (8) and (9), respectively, after repeating the recursions (10) until $i = k$. Since k is nondecreasing, one need only maintain the summary variables $W_i(b)$, $Y_i(b)$, and $Z_i(b)$.

To this end, one can use vectors $S[\cdot]$, $W[\cdot]$, $Y[\cdot]$, and $Z[\cdot]$, where the lth entry of each vector is used to store summary data for batch size b_l. When the first batch of size b_l is complete, the vectors are initialized as follows:

1. $s = S_{b_l}$.
2. $w = s$.
3. $Z[l] = 0$.
4. $W[l] = w$.
5. $S[l] = s$.
6. $Y[l] = W[l]^2$.

For every complete subsequent batch of size b_l, the vectors are updated by:

1. $w = s - S[l]$.
2. $Z[l] = Z[l] + w \cdot W[l]$.
3. $W[l] = w$.
4. $S[l] = s$.
5. $Y[l] = Y[l] + W[l]^2$.

A potential problem with the above updates is the lack of knowledge of the batch sizes that will be used. Hence, it may be necessary to maintain the vector entries corresponding to all potential batch sizes. Below we show how a large number of redundant operations can be eliminated.

Notice that (see Section 2.2) the set of potential batch sizes is $A_1 \cup A_2$, where $A_1 = \{2^l b_1, l = 0, 1, \ldots, \lfloor \log_2(\tilde{n}/(b_1 k_1)) \rfloor\}$, and $A_2 = \{2^l \tilde{b}_1, l \in I\}$, where

$$I = \begin{cases} \{0, \ldots, \lfloor \log_2(\tilde{n}/(\tilde{b}_1 \tilde{k}_1)) \rfloor\} \leftarrow \text{if } b_1 > 1,\ \tilde{n} \geq \tilde{b}_1 \tilde{k}_1 \\ \{1, \ldots, \lfloor \log_2(\tilde{n}/(\tilde{b}_1 \tilde{k}_1)) \rfloor\} \leftarrow \text{if } b_1 = 1,\ \tilde{n} \geq 2\tilde{b}_1 \tilde{k}_1 \\ \emptyset \leftarrow \quad \text{otherwise.} \end{cases}$$

Define the auxiliary variables

$$\begin{aligned} \theta_i(b) &= W_1(b)^2/2 + \sum_{j=2}^{\lceil i/2 \rceil} W_{2j-1}(b) W_{2j-2}(b), \\ \xi_i(b) &= \sum_{j=1}^{\lfloor i/2 \rfloor} W_{2j}(b) W_{2j-1}(b) \end{aligned} \tag{11}$$

and use (10) and (11) to obtain

$$Z_k(b) = -W_1(b)^2/2 + \theta_k(b) + \xi_k(b) \tag{12}$$

and

$$C_k(b) = \frac{-S_N^2/k + \theta_k(b) + \xi_k(b) + W_k(b)^2/2}{Y_k(b) - S_N^2/k}. \tag{13}$$

Notice that for odd $i \geq 3$

$$\begin{aligned} \theta_i(b) &= \theta_{i-2}(b) + W_i(b) W_{i-1}(b), \psi \\ \xi_i(b) &= \xi_{i-1}(b) \end{aligned} \tag{14}$$

and for even i

$$\begin{aligned} \xi_i(b) &= \xi_{i-2}(b) + W_i(b) W_{i-1}(b), \psi \\ \theta_i(b) &= \theta_{i-1}(b). \end{aligned} \tag{15}$$

Hence von Neumann's statistic can be computed from (13) after repeating (14) and (15) until $i = k$.

After some algebra, one can now show that for even k

$$Y_k(b/2) + 2\xi_k(b/2) = Y_{k/2}(b) \tag{16}$$

whereas for odd k

$$Y_k(b/2) - W_k(b/2)^2 + 2\xi_k(b/2) = Y_{(k-1)/2}(b). \tag{17}$$

The last two equalities show that, on each review, direct computation of $Y_k(b)$ is only required for the current batch size and the potential batch size resulting form the SQRT rule. This can be achieved by using two summary variables, y (for the current batch size) and $\tilde{y}$ (for the SQRT batch size). Whenever a batch size is eliminated, either (16) or (17) can be used to update the values of y or $\tilde{y}$ for the new batch size. Therefore, y, $\tilde{y}$, and the vectors $\theta[\cdot]$ and $\xi[\cdot]$ can be used to compute $Y_k(b)$ and $C_k(b)$.

When the first batch of size b_l is complete, the vectors are initialized as follows:

1. $w = s$.
2. $\theta[l] = w^2/2$.
3. $\xi[l] = 0$.
4. $W[l] = w$.
5. $S[l] = s$.

Upon completion of each subsequent batch of size b_l, these vectors are updated by:

1. $w = s - S[l]$.
2. If the number of batches is even,
   then $\xi[l] = \xi[l] + w \cdot W[l]$;
   otherwise, $\theta[l] = \theta[l] + w \cdot W[l]$.

3. $W[l] = w$.
4. $S[l] = S$.
5. $Y[l] = Y[l] + W[l]^2$.

The last issue that must be addressed is the testing for batch completion. Fortunately, the number of batch completion tests that are performed in either LBATCH of ABATCH can be reduced substantially by exploiting the special structure of the set of potential batch sizes.

The complexity of the two algorithms is derived as follows: The **Until** loop is executed no more than $\tilde{n}$ times. Thus, excluding the time for updating the vectors $S[\cdot]$, $Y[\cdot]$, $\theta[\cdot]$, and $\xi[\cdot]$, either algorithm takes $O(\tilde{n})$ time. Since the entries $\theta[l]$ are updated only after odd batch completions and the entries $\xi[l]$ are updated only after even batch completions, the total number of updates is bounded above by

$$\sum_{l=0}^{\lfloor \log_2(\tilde{n}/b_1)\rfloor} \frac{1}{2}\lfloor \tilde{n}/(2^l b_1)\rfloor + \sum_{l=0}^{\lfloor \log_2(\tilde{n}/\tilde{b}_1)\rfloor} \frac{1}{2}\lfloor \tilde{n}/(2^l \tilde{b}_1)\rfloor \le 2\tilde{n}.$$

Each update is performed in $O(1)$ time. Then the computational complexity of LBATCH and ABATCH is $O(\tilde{n}) = O(n)$.

Now observe that the vectors $S[\cdot]$, $Y[\cdot]$, $\theta[\cdot]$, and $\xi[\cdot]$ have $\max\{2\lfloor \log_2(\tilde{n}/b_1)\rfloor + 1, 2\lfloor \log_2(\tilde{n}/\tilde{b}_1)\rfloor + 2\} = O(\log_2 \tilde{n})$ entries. Hence both LBATCH and ABATCH require $O(\log_2 \tilde{n}) = O(\log_2 n)$ space.

Remark 1 The above numerical techniques can be extended to non-classical batch means methods with dynamic batching strategies. These methods include Overlapping Batch Means (Meketon and Schmeiser 1984) and Spaced Batch Means (Fox, Goldsman, and Swain 1990). The complexity issues are a problem under investigation.

REFERENCES

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., D. Goldsman, and B. Melamed. 1996. Large-sample results for batch means. To appear in *Management Science*.

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. 1978. Grouping observations in digital simulation. *Management Science* 24:510–521.

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

Fishman, G. S., and L. S. Yarberry. 1997. An implementation of the batch means method. To appear in *INFORMS Journal on Computing*.

Fox, B. L., D. Goldsman, and J. J. Swain. 1990. Spaced batch means. *Operations Research Letters* 10:255–266.

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

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.

Mechanic, H., and W. McKay. 1966. Confidence intervals for averages of dependent data in simulations II. Technical Report ASDD 17–202, IBM Corporation, Yorktown Heigths, New York.

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*. Boston: Birkhaüser.

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

Schriber, T. J., and R. W. Andrews. 1979. Interactive analysis of simulation output by the method of batch means. In *Proceedings of the 1979 Winter Simulation Conference*, eds. M. G. Spiegel, N. R. Nielsen, and H. J. Highland, 513–525. 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.

Yarberry, L. S. 1993. Incorporating a dynamic batch size selection mechanism in a fixed-sample-size batch means procedure. Ph.D. dissertation, Department of Operations Research, University of North Carolina, Chapel Hill, North Carolina.

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*.

GEORGE FISHMAN is Professor of Operations Research at the University of North Carolina at Chapel Hill. His areas of research include Monte Carlo methods and discrete-event simulation, and he has contributed to the literature on the design of simulation experiments, sample path data analysis, pseudorandom number generation and non-uniform variate generation. His most recent book is *Monte Carlo: Concepts, Algorithms, and Applications*, published by Springer-Verlag in 1996.

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 Program Chair for the 1994 conference.