Proceedings of the 2004 Winter Simulation Conference
R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

STEADY-STATE SIMULATION ANALYSIS USING ASAP3

Natalie M. Steiger

Maine Business School
University of Maine
Orono, ME 04469-5723, U.S.A.

Emily K. Lada

Statistical and Applied Mathematical Sciences Institute
Research Triangle Park, NC 27709-4006, U.S.A.

James R. Wilson
Jeffrey A. Joines

Colleges of Engineering and Textiles
North Carolina State University
Raleigh, NC 27695-7906, U.S.A.

Christos Alexopoulos
David Goldsman

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

ABSTRACT

We discuss ASAP3, a refinement of the batch means algorithms ASAP and ASAP2. ASAP3 is a sequential procedure designed to produce a confidence-interval estimator for the expected response of a steady-state simulation that satisfies user-specified precision and coverage-probability requirements. ASAP3 operates as follows: the batch size is increased until the batch means pass the Shapiro-Wilk test for multivariate normality; and then ASAP3 fits a first-order autoregressive (AR(1)) time series model to the batch means. If necessary, the batch size is further increased until the autoregressive parameter in the AR(1) model does not significantly exceed 0.8. Next ASAP3 computes the terms of an inverse Cornish-Fisher expansion for the classical batch means t-ratio based on the AR(1) parameter estimates; and finally ASAP3 delivers a correlation-adjusted confidence interval based on this expansion. ASAP3 compared favorably with other batch means procedures (namely, ABATCH, ASAP, ASAP2, and LBATCH) in an extensive experimental performance evaluation.

1 INTRODUCTION

In discrete-event simulation, we are often interested in estimating the steady-state mean μ_X of a stochastic output process $\{X_j : j = 1, 2, \ldots\}$ generated by a single, prolonged simulation run. Assuming the target process is stationary and given a time series of length n that is part of a single realization of this process, we see that a natural point estimator of μ_X is the sample mean, given by $\overline{X}(n) = n^{-1} \sum_{j=1}^{n} X_j$. We also require some indication of the precision of this point estimator; and typically we construct a confidence interval (CI) for μ_X with a user-specified probability $1 - \alpha$ of covering the point μ_X, where $0 < \alpha < 1$. The CI for μ_X should satisfy two criteria: (*i*) it is approximately valid—that is, its coverage probability is sufficiently close to the nominal level $1 - \alpha$; and (*ii*) it has sufficient precision—that is, it is narrow enough—to be meaningful in the context of the application at hand.

In the simulation analysis method of nonoverlapping batch means (NOBM), the sequence of simulation-generated outputs $\{X_j : j = 1, \ldots, n\}$ is divided into k adjacent nonoverlapping batches, each of size m. For simplicity, we assume that n is a multiple of m so that $n = km$. The sample mean for the jth batch is

$$Y_j(m) = \frac{1}{m} \sum_{i=m(j-1)+1}^{mj} X_i \quad \text{for} \quad j = 1, \ldots, k; \qquad (1)$$

and the grand mean of the individual batch means,

$$\overline{Y} = \overline{Y}(m, k) = \frac{1}{k} \sum_{j=1}^{k} Y_j(m), \qquad (2)$$

is used as a point estimator for μ_X (note that $\overline{Y}(m, k) = \overline{X}(n)$). We construct a CI estimator for μ_X that is centered on a point estimator like (2), where in practice we may exclude some initial batches to eliminate the effects of initialization bias.

If the batch size m is sufficiently large so that the batch means $\{Y_j(m) : j = 1, \ldots, k\}$ are approximately independent and identically distributed (i.i.d.) normal random variables with mean μ_X, then we can apply classical results concerning Student's t-distribution (see, for example, Alexopoulos and Goldsman 2004) to compute a confidence

interval for μ_X from the batch means. For this purpose we compute the sample variance of the k batch means for batches of size m,

$$S^2_{m,k} = \frac{1}{k-1}\sum_{j=1}^{k}\left[Y_j(m) - \overline{Y}(m,k)\right]^2.$$

If the original simulation-generated process $\{X_j : j = 1, \ldots, n\}$ is stationary and weakly dependent as specified, for example, in Theorem 1 of Steiger and Wilson (2001), then it follows that as $m \to \infty$ with k fixed so that $n \to \infty$, an asymptotically valid $100(1-\alpha)\%$ CI for μ_X is

$$\overline{Y}(m,k) \pm t_{1-\alpha/2,k-1}\frac{S_{m,k}}{\sqrt{k}}, \tag{3}$$

where $t_{1-\alpha/2,k-1}$ denotes the $1-\alpha/2$ quantile of Student's t-distribution with $k-1$ degrees of freedom.

Conventional NOBM procedures such as ABATCH and LBATCH (Fishman and Yarberry 1997, Fishman 1998) are based on (3); and they are designed to determine the batch size, m, and the number of batches, k, that are required to satisfy approximately the assumption of i.i.d. normal batch means. If this assumption is satisfied exactly, then we will obtain a CI whose actual coverage probability is exactly equal to the nominal level $1-\alpha$. By contrast, the more recent NOBM procedures ASAP (Steiger 1999; Steiger and Wilson 1999, 2000, 2002a, 2002b) and ASAP2 (Steiger et al. 2002) are designed to determine a batch size and an initial warm-up period sufficient to ensure that batch means computed beyond the warm-up period are approximately multivariate normal with identically distributed marginals (that is, they constitute a stationary Gaussian process) but are not necessarily independent. If the resulting batch means are correlated, then the classical NOBM t-ratio underlying (3) does not possess Student's t-distribution with $k-1$ degrees of freedom so that an appropriate modification of (3) is required to yield an approximately valid CI for μ_X.

Both ASAP and ASAP2 are designed to adjust (3) so as to account for any correlations among the batch means that those procedures finally deliver; and the required correlation adjustment is based on an inverse Cornish-Fisher expansion for the classical NOBM t-ratio. There is substantial experimental evidence that when ASAP or ASAP2 is applied with a user-specified absolute- or relative-precision requirement for the final delivered confidence interval, either procedure outperforms conventional NOBM procedures such as ABATCH and LBATCH in a large class of steady-state simulation models (Steiger and Wilson 2002a, Steiger et al. 2002). However, when either ASAP or ASAP2 is applied without a precision requirement, the delivered confidence intervals may exhibit excessive variability in some applications—that is, the variance and coefficient of variation of the CI half-lengths may be unacceptably large (Steiger and Wilson 2002a; Steiger et al. 2002; Lada, Wilson, and Steiger 2003).

In this paper we discuss ASAP3, a refinement of ASAP and ASAP2 that retains the advantages of its predecessors but is specifically designed to prevent excessive CI variability even in the absence of a precision requirement. Since the previously cited studies reveal that ABATCH and ASAP2 outperform LBATCH and ASAP respectively, in this paper we limit the our experimental performance evaluation to a comparison of ABATCH, ASAP2, and ASAP3. In §2 we provide an overview of ASAP3 and a formal algorithmic statement of the procedure. In §3 we summarize some of the results of our experimental performance evaluation; and in §4 we present our main conclusions. Full details on this work are available in Steiger et al. (2004).

2 OVERVIEW OF ASAP3

Figure 1 displays a high-level flow chart of ASAP3. The procedure operates as follows. The series of simulation outputs is divided initially into $k = 256$ batches, each of a user-specified size m (where the default initial batch size $m = 16$); and the corresponding batch means are computed as in (1). The first four batches are ignored to reduce the potential effects of initialization bias, and the remaining $k' = k - 4 = 252$ batch means are organized into adjacent nonoverlapping groups, where each group consists of four consecutive batch means. We select every other group of four consecutive batch means to form a sample of 32 four-dimensional vectors that we will test for stationary multivariate normality. If this test is failed, then the batch size m is increased by the factor $\sqrt{2}$; additional data are obtained; and the process of computing 256 batch means with the new batch size and testing for multivariate normality proceeds as outlined above using all accumulated data. ASAP3 iteratively performs this sequence of steps, systematically decreasing the significance level δ for the multivariate normality test on successive iterations until that test is finally passed. (See the last paragraph of this section and Steiger et al. 2004 for further explanation of this issue.)

Upon accepting the hypothesis of stationary multivariate normality of the batch means, we fit a first-order autoregressive (that is, AR(1)) time series model to the 252 batch means that remain after skipping the first group of four batch means. Adapting the notation of Box, Jenkins, and Reinsel (1994) to the nomenclature used here, we let $\left\{\widetilde{Y}_{j-4} \equiv Y_j(m) - \mu_X : j = 5, \ldots, k\right\}$ denote the corresponding deviations of the truncated batch means from the unknown steady-state mean μ_X. The ℓth observation of such an AR(1) process can be expressed as

$$\widetilde{Y}_\ell = \varphi\widetilde{Y}_{\ell-1} + a_\ell \quad \text{for} \quad \ell = 1, 2, \ldots, \tag{4}$$

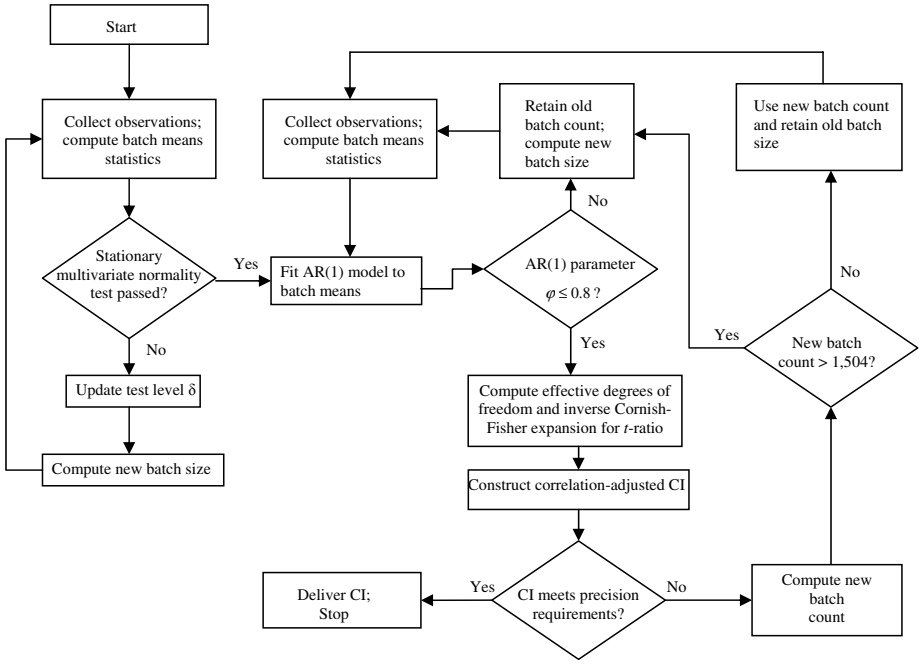

Figure 1: High-level Flow Chart of ASAP3

where the autoregressive parameter $\varphi \in (-1, 1)$ and $\{a_\ell\}$ are i.i.d. normal residuals with mean 0 and variance σ_a^2.

After fitting the AR(1) model (4) to the truncated batch means $\{Y_j(m) : j = 5, \ldots, k\}$, we apply a normalizing arc sine transformation to the autoregressive parameter estimator $\widehat{\varphi}$ so as to test the null hypothesis that the correlation between adjacent batch means (that is, φ) is at most 0.8 versus the alternative hypothesis that $\varphi > 0.8$. We have found that the condition $\varphi > 0.8$ is associated with excessive variability in the CIs delivered by ASAP and ASAP2. If the null hypothesis is rejected, then the batch count is retained; the batch size m is increased by a factor projected to reduce the lag-one correlation between batch means to the threshold 0.8 (see the last paragraph of this section and Steiger et al. 2004 for further explanation of this issue); additional data are obtained; and the process of computing batch means, fitting an AR(1) model, and testing the autoregressive parameter estimator proceeds as outlined above. ASAP3 iteratively performs the sequence of steps described in this paragraph until we finally obtain a batch size m for which we accept the null hypothesis of nonexcessive correlation between adjacent batch means.

Next ASAP3 constructs a CI for μ_X that has been adjusted to account for the remaining (nonexcessive) correlations between the k' batch means for batches of the current size m. The correlation adjustment uses an inverse Cornish-Fisher expansion (Stuart and Ord 1994) for the classical NOBM t-ratio

$$t = \left[\overline{Y}(m, k') - \mu_X \right] \Big/ \sqrt{S^2_{m,k'}/k'}; \tag{5}$$

and the terms of this expansion are computed from the parameter estimates $\widehat{\varphi}$ and $\widehat{\sigma}_a^2$ that are obtained by fitting the AR(1) model (4) to the current set of k' truncated batch means. Based on this approach, a correlation-adjusted $100(1-\alpha)\%$ CI for μ_X is

$$\overline{Y}(m, k') \pm \left[\left(\frac{1}{2} + \frac{\widehat{\kappa}_2}{2} - \frac{\widehat{\kappa}_4}{8} \right) z_{1-\alpha/2} + \frac{\widehat{\kappa}_4}{24} z^3_{1-\alpha/2} \right] \sqrt{\frac{\widehat{\mathrm{Var}}[Y(m)]}{k'}}, \tag{6}$$

where: $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution; $\widehat{\kappa}_2$ and $\widehat{\kappa}_4$ respectively denote estimators of the second and fourth cumulants of the t-ratio (5); $\widehat{\mathrm{Var}}[Y(m)]$ denotes an estimator of the variance of the batch means; and the statistics $\widehat{\kappa}_2$, $\widehat{\kappa}_4$, and $\widehat{\mathrm{Var}}[Y(m)]$ are computed from $\widehat{\varphi}$ and $\widehat{\sigma}_a^2$ as detailed in Steiger et al. (2004).

If additional observations of the target process must be generated by the user's simulation model before a CI can be delivered that has the form (6) and the required precision, then ASAP3 estimates a new, larger sample size based on the ratio of the current iteration's CI half-length to the desired CI half-length (see the last paragraph of this section and Steiger et al. 2004 for further explanation of this issue). Then ASAP3 must be called again with the additional data; and this cycle of simulation followed by analysis may be repeated several times before ASAP3 finally delivers a CI with the required precision.

Subsequent iterations of ASAP3 that are performed to satisfy the user-specified precision requirement do not repeat the test of the overall set of batch means for stationary

multivariate normality; but on every iteration of ASAP3, we fit an AR(1) process to the latest set of batch means, test the hypothesis that $\varphi \leq 0.8$, and if necessary increase the batch size by the updated factor that is currently projected to reduce the lag-one correlation between batch means to the threshold 0.8. Thus each additional iteration of ASAP3 that is performed solely to satisfy the precision requirement will involve the following operations: (*i*) obtaining additional simulation-generated data; (*ii*) recomputing the batch means with a new batch size or computing additional batch means of the same size; (*iii*) retesting the hypothesis that $\varphi \leq 0.8$ with progressively larger batch sizes until that hypothesis is accepted; and (*iv*) reconstructing the CI for μ_X and testing that CI for conformance to the user's precision requirement, if necessary computing the total sample size required for the next iteration of ASAP3. Successive iterations of ASAP3 involving operations (*i*)–(*iv*) above are performed until the precision requirement is met.

ASAP3 requires the following user-supplied inputs:

- a simulation-generated output process $\{X_j : j = 1, \ldots, n\}$ from which the steady-state expected response μ_X is to be estimated;
- the desired CI coverage probability $1 - \alpha$, where $0 < \alpha < 1$; and
- an absolute or relative precision requirement specifying the final confidence-interval half-length in terms of (*i*) a maximum acceptable half-length H^* (for an absolute precision requirement); or (*ii*) a maximum acceptable fraction r^* of the magnitude of the CI midpoint (for a relative precision requirement).

ASAP3 delivers the following outputs:

- a nominal $100(1 - \alpha)\%$ CI for μ_X that satisfies the specified absolute or relative precision requirement, provided no additional simulation-generated observations are required; or
- a larger total sample size n to be supplied to ASAP3 when it is executed again.

A formal algorithmic statement of ASAP3 is displayed in Figure 2. Note that in Figure 2, if a and b are given constants with $a < b$, then we take

$$\text{mid}(a, x, b) \equiv \begin{cases} a, & \text{if } x < a, \\ x, & \text{if } a \leq x \leq b, \\ b, & \text{if } x > b, \end{cases} \quad \text{and } x_+ \equiv \max\{0, x\}.$$

Three points about Figure 2 warrant special comment.

- On the ith iteration of the multivariate normality test in step **[2]**, we set δ_i, the significance level of the test, so as to control the overall level of type I error and avoid explosive growth of the overall sample size required for this step of ASAP3.
- In step **[3]**, we estimate the batch-size multiplier that is required to reduce the lag-one correlation between batch means to the threshold 0.8.
- In step **[5]**, we estimate the increase in the number of batches or the increase in the batch size that is needed to satisfy the precision requirement.

Steiger et al. (2004) present a complete development of the steps of the ASAP3. A stand-alone Windows-based version of ASAP3 and a user's manual are available online via Steiger et al. (2003).

3 EXPERIMENTAL PERFORMANCE EVALUATION

To evaluate the performance of ASAP3 with respect to the coverage probability of its CIs, the mean and variance of the half-length of its CIs, and its total sample size, we applied ASAP3 together with ABATCH and ASAP2 to a large suite of test problems. The experimental design includes some problems typically used to test simulation output analysis procedures and some problems more closely resembling real-world applications. To demonstrate the robustness of ASAP3, we limit our discussion here to an $M/M/1$ queue waiting time process for a system with an empty-and-idle initial condition, an interarrival rate of 0.9, and a service rate of 1.0. In this system the steady-state server utilization is 0.9 and the steady-state expected waiting time in the queue is $\mu_X = 9$.

The $M/M/1$ queue waiting time process is a particularly difficult test problem for several reasons: (*i*) the magnitude of the initialization bias is substantial and decays relatively slowly; (*ii*) in steady-state operation the autocorrelation function of the waiting time process decays very slowly with increasing lags; and (*iii*) in steady-state operation the marginal distribution of waiting times has an exponential tail and is therefore markedly nonnormal. Because of these characteristics, we can expect slow convergence to the classical requirement that the batch means are i.i.d. normal. This test problem clearly reveals one of the principal advantages of the ASAP3 algorithm—namely, that ASAP3 does not rely on any test for independence of the batch means. The steady-state mean response is available analytically for this test problem; thus we were able to evaluate the performance of ABATCH, ASAP2, and ASAP3 in terms of actual versus nominal coverage probabilities for the CIs delivered by each of these procedures. Experimental results for the other test problems may be found in Steiger et al. (2004).

Our experiments included 400 independent replications of each batch means procedure to construct nominal 90% and 95% CIs that satisfied four different precision requirements. For the case of no precision requirement, we took $r^* = 0$

[0] Set iteration index $i \leftarrow 1$, $m_1 \leftarrow$ user-specified initial batch size (default = 16), initial batch count $k_1 \leftarrow 256$, initial sample size $n_1 \leftarrow k_1 m_1$ with $n_0 \leftarrow 0$, truncated initial batch count $k_1' \leftarrow k_1 - 4$, $1 - \alpha \leftarrow$ user-specified CI coverage probability (default = 0.90), size of test for autoregressive parameter $\alpha_{\text{arp}} \leftarrow 0.01$, initial size of test for stationary multivariate normality $\delta_1 \leftarrow 0.1$ with parameter $\omega \leftarrow 0.18421$ controlling the test size in step **[2]** on subsequent iterations, and indicator that normality test was passed `MVTestPassed` $\leftarrow$ 'no';

if a relative precision requirement is given, **then** set `RelPrec` $\leftarrow$ 'yes' and $r^* \leftarrow$ the user-specified fraction of the magnitude of the CI midpoint that defines the maximum acceptable CI half-length;

if an absolute precision requirement is given, **then** set `RelPrec` $\leftarrow$ 'no' and $H^* \leftarrow$ the user-specified maximum acceptable CI half-length;

if no precision level is specified **then** set `RelPrec` $\leftarrow$ 'no', $r^* \leftarrow 0$, and $H^* \leftarrow 0$.

[1] Start (or restart) the simulation to generate the data $\{X_j : j = n_{i-1} + 1, \ldots, n_i\}$ required for the current iteration i;

Compute the k_i batch means $\left\{Y_j(m_i) : j = 1, \ldots, k_i\right\}$; and after skipping the initial spacer $\{Y_1(m_i), Y_2(m_i), Y_3(m_i), Y_4(m_i)\}$, compute the truncated grand mean,

$$\overline{Y}(m_i, k_i') \leftarrow \frac{1}{k_i' m_i} \sum_{\ell=4m_i+1}^{n_i} X_\ell = \frac{1}{k_i'} \sum_{j=5}^{k_i} Y_j(m_i); \tag{7}$$

if `MVTestPassed`='yes', **then goto [3]**.

[2] From the truncated batch means $\left\{Y_j(m_i) : j = 5, \ldots, k_i\right\}$, select every other group of four successive batch means to build the 4×1 vectors

$$\left\{\mathbf{y}_\ell = \left[Y_{5+(\ell-1)8}(m_i), Y_{6+(\ell-1)8}(m_i), Y_{7+(\ell-1)8}(m_i), Y_{8+(\ell-1)8}(m_i)\right]^{\mathrm{T}} : \ell = 1, \ldots, 32\right\};$$

To test the hypothesis

$$\mathcal{H}_{\text{mvn}}: \left\{\mathbf{y}_\ell : \ell = 1, \ldots, 32\right\} \text{ are i.i.d. four-dimensional normal random vectors,}$$

evaluate $\delta_i = \delta_1 \exp\left[-\omega(i-1)^2\right]$, the significance level for the test, and W_i^*, the multivariate Shapiro-Wilk statistic computed from the $\{\mathbf{y}_\ell\}$ according to equations (10)–(12) of Steiger et al. (2004);

if $W_i^* < w_{\delta_i}^*$, the δ_i quantile of the distribution of W_i^* under the null hypothesis $\mathcal{H}_{\text{mvn}}$, so that $\mathcal{H}_{\text{mvn}}$ is rejected at significance level δ_i, **then**
set $i \leftarrow i + 1$, $k_i \leftarrow 256$, $k_i' \leftarrow k_i - 4$, $m_i \leftarrow \left\lfloor \sqrt{2} m_{i-1} \right\rfloor$, and $n_i \leftarrow k_i m_i$;
goto [1];
else
set `MVTestPassed` $\leftarrow$ 'yes';
goto [3].

Figure 2: Algorithmic Statement of ASAP3

[3] Fit an AR(1) model (4) to the truncated batch means $\{Y_j(m_i) : j = 5, \ldots, k_i\}$ so as to obtain the estimator $\widehat{\varphi}$ of the autoregressive parameter φ;

Test the hypothesis $\mathcal{H}_{\text{arp}} : \varphi \le 0.8$ at the level of significance α_{arp} by checking for the condition

$$\widehat{\varphi} \le \sin\Big(0.927 - z_{1-\alpha_{\text{arp}}} \Big/ \sqrt{k_i'} \Big) ; \tag{8}$$

if $\mathcal{H}_{\text{arp}}$ is rejected at significance level α_{arp}, **then**

set $\theta \leftarrow \text{mid}\Big\{ \sqrt{2},\ \ln\Big[\sin\Big(0.927 - z_{1-\alpha_{\text{arp}}} \Big/ \sqrt{k_i'} \Big) \Big] \Big/ \ln(\widehat{\varphi}),\ 4 \Big\}$,

$i \leftarrow i+1, \quad k_i \leftarrow k_{i-1}, \ k_i' \leftarrow k_i - 4, \quad m_i \leftarrow \lceil \theta m_{i-1} \rceil, \quad$ and $\quad n_i \leftarrow k_i m_i$; **goto [1]**;

else

goto [4].

[4] Using the estimators $\widehat{\varphi}$ and $\widehat{\sigma}_a^2$ for the AR(1) model (4), compute $\widehat{\text{Var}}[Y(m_i)]$ and $\widehat{\text{Var}}\big[\, \overline{Y}(m_i, k_i') \,\big]$ from equations (15)–(16) of Steiger et al. (2004);

For the NOBM t-ratio (5), compute the estimated effective degrees of freedom $\widehat{\nu}_{\text{eff}}$ from equation (33) of Steiger et al. (2004);

Compute $\widehat{\kappa}_2$ and $\widehat{\kappa}_4$, the estimators, respectively, of the second and fourth cumulants of the t-ratio (5), by inserting $\widehat{\text{Var}}[Y(m_i)]$, $\widehat{\text{Var}}\big[\, \overline{Y}(m_i, k_i') \,\big]$, and $\widehat{\nu}_{\text{eff}}$ into the computing expressions for κ_2 and κ_4 given in equations (31)–(32) of Steiger et al. (2004);

Calculate the half-length of the correlation-adjusted CI,

$$H \leftarrow \left[\left(\frac{1}{2} + \frac{\widehat{\kappa}_2}{2} - \frac{\widehat{\kappa}_4}{8} \right) z_{1-\alpha/2} + \frac{\widehat{\kappa}_4}{24} z_{1-\alpha/2}^3 \right] \sqrt{\frac{\widehat{\text{Var}}[Y(m_i)]}{k_i'}} \ ;$$

Construct the correlation-adjusted CI,

$$\overline{Y}(m_i, k_i') \pm H. \tag{9}$$

[5] **if** `RelPrec`='yes' **then** set $H^* \leftarrow r^* \big| \overline{Y}(m_i, k_i') \big|$;

if $(H \le H^*)$ **or** $(r^* = 0$ **and** $H^* = 0)$, **then**

deliver $\overline{Y}(m_i, k_i') \pm H$ and **stop**;

else

Estimate additional batches needed to satisfy the precision requirement,

$$k'' = \max\Big\{ \Big\lceil \big(H \big/ H^* \big)^2 k_i' \Big\rceil - k_i',\ 1 \Big\} ;$$

If $k_i + k'' \le 1{,}504$, **then**

set $i \leftarrow i+1, \quad k_i \leftarrow k_{i-1} + k'', \ k_i' \leftarrow k_i - 4, \quad m_i \leftarrow m_{i-1}, \quad$ and $\quad n_i \leftarrow m_i k_i$; **goto [1]**;

else

Find the root θ of the equation $\quad \theta \left(1 - \widehat{\varphi}_+^{\theta} \right)^2 = \big(H \big/ H^* \big)^2 \left(1 - \widehat{\varphi}_+ \right)^2$,

set $\theta \leftarrow \text{mid}\Big(\sqrt{2},\ \theta,\ 4 \Big), \quad i \leftarrow i+1, \quad k_i \leftarrow k_{i-1}, \quad k_i' \leftarrow k_i - 4,$

$m_i \leftarrow \lceil \theta m_{i-1} \rceil$, $\quad$ and $\quad n_i \leftarrow m_i k_i$; **goto [1]**.

Figure 2 (Continued): Algorithmic Statement of ASAP3

and $H^* = 0$ in the initialization step **[0]** of Figure 2 so that we continued the simulation of each test problem until ASAP3 completed the following operations: step **[2]** (so that the batch means passed the test for stationary multivariate normality); step **[3]** (so that the batch means passed the test for acceptable lag-one correlation); step **[4]** (so that the first CI of the form (9) could be constructed); and finally step **[5]** (so that ASAP3 stopped immediately, delivering the first CI of the form (9) that was constructed). For the cases of the precision requirements $\pm 15\%$, $\pm 7.5\%$, and $\pm 3.75\%$, we continued the simulation of each test problem until ASAP3 delivered a CI of the form (9) that satisfied the stopping criterion in step **[5]** with $r^* = 0.15$, 0.075, and 0.0375, respectively.

In addition to the experimentation using the ASAP3 algorithm, we performed 400 independent replications of the ASAP2 algorithm under the same precision requirements described above. Recall that unlike ASAP3, ASAP2 does not include the test (8) for acceptable correlation between adjacent batch means.

Since ABATCH lacks a method for determining sample size, we passed to this procedure the same data sets used by ASAP3. Based on all our computational experience with ASAP2 and ASAP3, we believe that the results given below are typical of the performance of ASAP2 and ASAP3 that can be expected in many practical applications. On the other hand, ABATCH is a nonsequential procedure whose proper operation may require direct user intervention (Fishman 1998); and thus it is not clear that the following results exemplify the performance of ABATCH in practical applications. Nevertheless, we believe that the results given below provide some basis for comparing the performance of ABATCH, ASAP2, and ASAP3.

Since each CI was replicated 400 times, the standard error of the coverage estimator for CIs with nominal 90% coverage probability is approximately 1.5%; and for CIs with nominal 95% coverage probability, the standard error of the coverage estimator is approximately 1.1%. As explained below, this level of precision in the estimation of coverage probabilities turned out to be sufficient to reveal significant differences in the performance of ASAP3 compared with that of ASAP2 and ABATCH in the test problems presented here.

Table 1 summarizes the experimental performance of the procedures ABATCH, ASAP2, and ASAP3 when they were applied to the waiting times in the $M/M/1$ queue. As can be seen from this table, ASAP3 outperformed ABATCH with respect to CI coverage for the first three precision requirements. As we demanded improved levels of precision, we were of course forced to perform more sampling. For the precision requirement of $\pm 3.75\%$, the three algorithms gave similar results. The results in Table 1 suggest that ABATCH will give satisfactory coverage if this procedure is supplied with an adequate amount of data; however, ABATCH provides no mechanism for determining the amount of data that should be used. No average sample sizes are shown in the table for the ABATCH procedure since on each replication of ASAP3 and ABATCH, these two procedures used exactly the same data set, whose size was determined by the stopping rule in step **[5]** of ASAP3. Table 2 of Steiger and Wilson (2002a) shows that simply adding an absolute- or relative-precision stopping rule to ABATCH will not generally yield acceptable performance for this procedure. A desirable feature of ASAP3 is that it usually determines a sample size sufficient to yield acceptable results.

In the absence of a precision requirement, ASAP2-generated confidence intervals were highly variable in their half-lengths. Imposing the requirement (8) that the lag-one correlation between the batch means must not significantly exceed 0.8 greatly reduced the variability of the half-lengths of the CIs generated by ASAP3, as shown in Table 1. Moreover, in terms of CI coverage, ASAP3 performed as well as ASAP2 in the no precision case.

4 CONCLUSIONS

The undercoverage problem encountered with ASAP was virtually eliminated by removal of the test for independence of the batch means. Both ASAP2 and ASAP3 test only for stationary multivariate normality of the batch means and always deliver a CI adjusted for correlation, if any, among the final batch means. Excessive variabilities seen with ASAP in the final sample sizes, and to some extent in the final CI half-lengths, were partially resolved in ASAP2 by decreasing the significance level of the test for stationary multivariate normality on each iteration of that test. Moreover, the means and variances of the final CI half-lengths delivered by ASAP3 were greatly reduced in comparison with the corresponding quantities delivered by ASAP and ASAP2; and ASAP3 has achieved this performance improvement by progressively increasing the batch size until we can conclude that the correlation between adjacent batch means does not significantly exceed 0.8 in the sense that a one-sided upper 99% confidence interval for this correlation lies entirely below 0.8.

ASAP3 is primarily designed for use in conjunction with a user-specified absolute or relative precision requirement on the final CI; and when it is used in this way, ASAP3 generally delivers CIs whose coverage probability is close to the nominal level. On the basis of all the experimentation we have performed with the procedure, ASAP3 appears to deliver CIs whose coverage probability is reasonably close to the nominal level even in the absence of a precision requirement; but in such cases there is of course no guarantee that the resulting CIs will be narrow enough to be useful in practice. Although ASAP3 does not provide a definitive resolution of all problems associated with the batch means method for steady-state simulation output analysis, many

Table 1: Performance of Batch Means Procedures for the $M/M/1$ Queue Waiting Time Process with Traffic Intensity $\tau = 0.9$ Based on 400 Independent Replications of Nominal 90% and 95% Confidence Intervals

Precision Requirement	Nominal 90% CIs ABATCH	ASAP3	ASAP2	Nominal 95% CIs ABATCH	ASAP3	ASAP2
NO PRECISION						
avg. sample size		31,181	22,554		31,181	22,554
coverage	76.0%	87.5%	88.0%	81.8%	91.5%	90.3%
avg. rel. precision	0.161	0.239	0.579	0.193	0.290	0.730
avg. CI half length	1.388	2.072	6.440	1.669	2.521	8.300
var. CI half length	0.112	0.348	167.000	0.164	0.535	350.000
±15% PRECISION						
avg. sample size		103,742	93,374		140,052	126,839
coverage	80.5%	91.0%	90.0%	87.8%	95.5%	94.5%
avg. rel. precision	0.098	0.134	0.135	0.104	0.136	0.136
avg. CI half length	0.865	1.182	1.184	0.921	1.206	1.204
var. CI half length	0.020	0.026	0.025	0.023	0.020	0.020
±7.5% PRECISION						
avg. sample size		287,568	281,022		382,958	382,040
coverage	85.8%	89.5%	92.0%	92.3%	94.3%	96.0%
avg. rel. precision	0.063	0.070	0.070	0.066	0.071	0.071
avg. CI half length	0.561	0.627	0.628	0.588	0.632	0.633
var. CI half length	0.005	0.002	0.002	0.005	0.002	0.002
±3.75% PRECISION						
avg. sample size		969,011	943,498		1,341,522	1,331,887
coverage	88.8%	89.5%	92.0%	93.3%	93.5%	95.5%
avg. rel. precision	0.035	0.036	0.036	0.036	0.036	0.036
avg. CI half length	0.318	0.320	0.323	0.323	0.321	0.322
var. CI half length	0.001	4.4E-4	3.0E-4	0.001	3.8E-4	3.0E-4

of the undesirable behaviors of its predecessors ASAP and ASAP2 have been eliminated; and there is good evidence to show that ASAP3's performance in practice compares favorably with other well-known batch means procedures. We believe the basic approach of ASAP3 has the potential to lead to new developments in the method of batch means.

Additional experimental results, follow-up papers and revised software, will be available on the website <www.ie.ncsu.edu/jwilson>.

REFERENCES

Alexopoulos, C., and D. Goldsman. 2004. To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation* 14 (1): 76–114.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*. 3rd ed. Englewood Cliffs, New Jersey: Prentice Hall.

Fishman, G. S. 1998. LABATCH.2: Software for statistical analysis of simulation sample path data. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 131–139. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <www.informs-cs.org/wsc98papers/017.PDF> [accessed April 11, 2004].

Fishman, G. S., and L. S. Yarberry. 1997. An implementation of the batch means method. *INFORMS Journal on Computing* 9 (3): 296–310.

Lada, E. K., J. R. Wilson, and N. M. Steiger. 2003. A wavelet-based spectral method for steady-state simulation analysis. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 422–430. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <www.informs-cs.org/wsc03papers/052.pdf> [accessed April 11, 2004].

Steiger, Natalie Miller. 1999. Improved batching for confidence interval construction in steady-state simulation. Doctoral dissertation, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available online via <www.lib.ncsu.edu/etd/public/etd-19231992992670/etd.pdf> [accessed April 11, 2004].

Steiger, N. M., E. K. Lada, J. R. Wilson, C. Alexopoulos, D. Goldsman, and F. Zouaoui. 2002. ASAP2: An improved batch means procedure for simulation output analysis. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 336–344. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via

`<www.informs-cs.org/wsc02papers/043.pdf>` [accessed April 11, 2004].

Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2003. ASAP3 software and user's manual. Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available online via `<ftp.ncsu.edu/pub/eos/pub/jwilson/installasap3.exe>` [accessed April 11, 2004].

Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2004. ASAP3: A batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation* to appear. Available online via `<ftp.eos.ncsu.edu/pub/jwilson/2003-34steiger.pdf>` [accessed August 16, 2004].

Steiger, N. M., and J. R. Wilson. 1999. Improved batching for confidence interval construction in steady-state simulation. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 442–451. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via `<www.informs-cs.org/wsc99papers/061.PDF>` [accessed April 11, 2004].

Steiger, N. M., and J. R. Wilson. 2000. Experimental performance evaluation of batch-means procedures for simulation output analysis. In *Proceedings of the 2000 Winter Simulation Conference*, ed. R. R. Barton, J. A. Joines, K. Kang, and P. A. Fishwick, 627–636. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via `<www.informs-cs.org/wsc00papers/084.PDF>` [accessed April 11, 2004].

Steiger, N. M., and J. R. Wilson. 2001. Convergence properties of the batch means method for simulation output analysis. *INFORMS Journal on Computing* 13 (4): 277–293.

Steiger, N. M., and J. R. Wilson. 2002a. An improved batch means procedure for simulation output analysis. *Management Science* 48 (12): 1569–1586.

Steiger, N. M., and J. R. Wilson. 2002b. ASAP software and user's manual. Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available online via `<ftp.ncsu.edu/pub/eos/pub/jwilson/installasap.exe>` [accessed April 11, 2004].

Stuart, A., and J. K. Ord. 1994. *Kendall's advanced theory of statistics, Volume 1: Distribution theory*. 6th ed. London: Edward Arnold.

AUTHOR BIOGRAPHIES

NATALIE M. STEIGER is an assistant professor of production and operations management in the University of Maine Business School. She is a member of IIE and INFORMS. Her e-mail address is `<nsteiger@maine.edu>`.

EMILY K. LADA is a postdoctoral fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI). She is a member of IIE and INFORMS. Her e-mail address is `<eklada@eos.ncsu.edu>`.

JAMES R. WILSON is professor and head of the Department of Industrial Engineering at North Carolina State University. He is a member of AAUW, ACM, ASA and INFORMS, and he is a Fellow of IIE. His e-mail address is `<jwilson@eos.ncsu.edu>`, and his web page is `<www.ie.ncsu.edu/jwilson>`.

JEFFREY A. JOINES is assistant professor in the Department of Textile Engineering, Chemistry, and Science at North Carolina State University. His email address is `<JeffJoines@ncsu.edu>`.

CHRISTOS ALEXOPOULOS is an associate professor in the School of Industrial & Systems Engineering at Georgia Tech. His e-mail address is `<christos@isye.gatech.edu>`, and his web page is `<www.isye.gatech.edu/~christos>`.

DAVID GOLDSMAN is a professor in the School of Industrial & Systems Engineering at Georgia Tech. His e-mail address is `<sman@isye.gatech.edu>`, and his web page is `<www.isye.gatech.edu/~sman>`.