

## A SEQUENTIAL PROCEDURE FOR ESTIMATING THE STEADY-STATE MEAN USING STANDARDIZED TIME SERIES

Christos Alexopoulos  
David Goldsman  
Peng Tang

H. Milton Stewart School of Industrial and Systems Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332-0205, USA

James R. Wilson

Edward P. Fitts Department of Industrial and Systems Engineering  
North Carolina State University  
Raleigh, NC 27695-7906, USA

### ABSTRACT

We propose SPSTS, an automated sequential procedure for computing point and confidence-interval (CI) estimators for the steady-state mean of a simulation output process. This procedure is based on variance estimators computed from standardized time series, and it is characterized by its simplicity relative to methods based on batch means and its ability to deliver CIs for the variance parameter of the output process. The effectiveness of SPSTS is evaluated via comparisons with methods based on batch means. In preliminary experimentation with the steady-state waiting-time process for the  $M/M/1$  queue with a server utilization of 90%, we found that SPSTS performed comparatively well in terms of its average required sample size as well as the coverage and average half-length of its delivered CIs.

### 1 INTRODUCTION

Simulation is perhaps the most widely used tool in the fields of industrial engineering, operations research, and the management sciences. Steady-state simulations play a fundamental role in system design, and they are particularly appropriate for evaluating long-run system performance or risk. For instance, what is the steady-state expected return from a certain financial management strategy, or what is the long-term probability of a default? Although there are now many commercial and public-domain software packages supporting the development of valid and efficient simulation models for complex systems, rarely have these packages been equipped with *comprehensive* facilities for performing rigorous, state-of-the-art statistical analysis of the outputs arising from steady-state simulation experiments. In many large-scale simulation applications, most of the effort is devoted to the development and execution of computer-based models, while relatively little attention is devoted to careful follow-up analysis of the final results. For questions about risk in the context of steady-state simulation analysis, there is not even much literature on the supporting theory, not to mention the lack of implementation of that theory in practical problems.

A fundamental problem in simulation output analysis concerns the computation of point and confidence interval (CI) estimators for the mean  $\mu$  of a stationary discrete-time stochastic process  $X \equiv \{X_i : i = 1, 2, \dots\}$ . The point estimation of  $\mu$  is actually a relatively easy problem when the underlying system starts in steady state; based on a simulation-generated time series  $\{X_1, \dots, X_n\}$  of length  $n$ , the sample mean  $\bar{X}_n \equiv n^{-1} \sum_{i=1}^n X_i$  is an unbiased estimator of  $\mu$ . To provide a measure of the sample mean's precision, an estimate of  $\text{Var}[\bar{X}_n]$

also needs to be calculated. If the  $X_i$  are independent and identically distributed (i.i.d.) random variables, then the sample variance  $S_n^2 \equiv (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  is an unbiased estimator of the marginal variance  $\sigma_X^2 = \text{Var}[X_1]$ , and  $\text{Var}[\bar{X}_n]$  can be estimated by  $S_n^2/n$ . Unfortunately, the  $X_i$  encountered in simulation output are typically correlated. While  $\bar{X}_n$  remains an unbiased estimator for  $\mu$ , the sample variance  $S_n^2$  can be a severely biased estimator for  $\sigma_X^2$ . In fact, if the autocovariance function  $R_j \equiv \text{Cov}[X_1, X_{1+j}]$  for  $j = 0, \pm 1, \pm 2, \dots$  is positive, then in many practical applications, one has  $E[S_n^2/n] \ll \text{Var}[\bar{X}_n]$  (Law 2007), and CI estimation is difficult. However, in such situations, valid CIs for  $\mu$  can still be obtained based on good estimators of the quantities  $\{\sigma_n^2 \equiv n\text{Var}[\bar{X}_n] : n = 1, 2, \dots\}$  or their limit  $\sigma^2 \equiv \lim_{n \rightarrow \infty} \sigma_n^2$ , which is called the (asymptotic) variance parameter of the process—provided that  $X$  is a process with short-range dependence so that  $\sigma^2$  is well defined and finite.

The simulation literature contains many techniques for estimating the quantities  $\sigma_n^2$  and  $\sigma^2$ , such as: nonoverlapping batch means (NBM) (Fishman 2001, Fishman and Yarberry 1997, Tafazzoli and Wilson 2011); overlapping batch means (OBM) (Meketon and Schmeiser 1984); and standardized time series (STS) (Schruben 1983; Alexopoulos et al. 2007b). Most of these techniques group observations into nonoverlapping or overlapping batches. A method relying on nonoverlapping batches typically divides the data into adjacent disjoint batches of equal size, calculates a corresponding estimator from each batch separately, and then forms an overall variance estimator by averaging the estimators from the various batches. Nonoverlapping batches are used with the following estimators for  $\sigma^2$ : NBM; STS area; STS Cramér–von Mises (CvM) (Alexopoulos et al. 2007b); folded area; and folded CvM (Antonini et al. 2009). The NBM estimator is simply the product of the batch size and the sample variance of the batch means. For STS area and CvM estimators, one forms an STS from each batch, computes an appropriate functional of each STS, and averages these quantities.

This paper is a step towards the development of effective sequential procedures for computing valid CIs for the steady-state mean of a simulation-generated process using variance estimators based on standardized time series. As for the remainder of the paper, Section 2 presents a review of the necessary methodology, Section 3 details SPSTS, the first of a new class of sequential procedures for steady-state simulation output analysis; Section 4 illustrates the performance of SPSTS in a test process that is often used as a benchmark problem for evaluating output-analysis procedures, and Section 5 summarizes our conclusions and recommendations for follow-up work.

## 2 PRELIMINARIES

The following discussion is necessary to establish basic notation and motivation for the estimators under study herein. Among the assumptions enumerated below, the first property is satisfied by a variety of stationary processes, including strongly mixing processes, associated stationary processes, and regenerating processes (Durrett 2005, Glynn and Iglehart 1990).

### 2.1 Assumptions and Relevant Variance Estimators

#### Assumptions A

**A.1** [Functional Central Limit Theorem (FCLT)] Assume  $\sum_{j=-\infty}^{\infty} j^2 |R_j| < \infty$  so that we have  $\sigma^2 = \lim_{n \rightarrow \infty} n\text{Var}[\bar{X}_n] = \sum_{j=-\infty}^{\infty} R_j < \infty$ . Further, assume  $\sigma^2 > 0$ , and that the sequence of random functions

$$Y_n(t) \equiv \frac{\lfloor nt \rfloor (\bar{X}_{\lfloor nt \rfloor} - \mu)}{\sigma \sqrt{n}} \quad \text{for } t \in [0, 1] \text{ and } n = 1, 2, \dots$$

satisfies  $Y_n(\cdot) \xrightarrow[n \rightarrow \infty]{\Rightarrow} \mathcal{W}(\cdot)$ , where:  $\lfloor \cdot \rfloor$  is the greatest integer function;  $\mathcal{W}(\cdot)$  is a standard Brownian motion process on  $[0, 1]$ ; and  $\xrightarrow[n \rightarrow \infty]{\Rightarrow}$  denotes weak convergence (as  $n \rightarrow \infty$ ) in the Skorohod space  $D[0, 1]$  of real-valued functions on  $[0, 1]$  that are right-continuous with left-hand limits.

**A.2** The weight function  $f(\cdot)$  on  $[0, 1]$  has a continuous second derivative and is normalized so that  $\int_0^1 \int_0^1 f(s)f(t)[\min(s,t) - st] ds dt = 1$ . We define  $F(t) \equiv \int_0^t f(s) ds$  and  $\bar{F}(t) \equiv \int_0^t F(s) ds$  for  $t \in [0, 1]$ ; and we let  $F \equiv F(1)$  and  $\bar{F} \equiv \bar{F}(1)$ .

Assumption A.1 implies that if  $\bar{X}_n$  and  $\hat{\sigma}^2$  are asymptotically independent as  $n \rightarrow \infty$  and if  $\hat{\sigma}^2 \xrightarrow[n \rightarrow \infty]{\text{d}} \sigma^2 \chi_v^2/v$ , where  $\chi_v^2$  denotes a chi-squared distribution with  $v$  degrees of freedom, then

$$\bar{X}_n \pm t_{v, 1-\alpha/2} \sqrt{\hat{\sigma}^2/n}$$

is an asymptotically valid  $100(1 - \alpha)\%$  CI for  $\mu$ , where  $t_{v, \beta}$  is the  $\beta$  quantile of Student's  $t$  distribution with  $v$  degrees of freedom depending on the type of the estimator. Let the unknown constant  $\gamma_1 \equiv 2 \sum_{j=1}^{\infty} jR_j$  as in Song and Schmeiser (1995). We use the following notation: (i)  $p(n) = O(q(n))$  means that there are positive constants  $c$  and  $n_0$  such that  $0 \leq p(n) \leq cq(n)$  for all  $n \geq n_0$ ; and (ii)  $p(n) = o(q(n))$  means that  $\lim_{n \rightarrow \infty} p(n)/q(n) = 0$ .

## 2.2 Estimators Based on Nonoverlapping Batches

The concept of batching has a long history in the simulation output analysis literature. Suppose we form  $b$  nonoverlapping batches, each consisting of  $m$  observations (so that  $n = bm$ ). Specifically, batch  $i$  consists of  $\{X_{(i-1)m+j} : j = 1, \dots, m\}$ . Finally, for  $i = 1, \dots, b$  and  $k = 1, \dots, m$ , let  $\bar{X}_{i,k} \equiv k^{-1} \sum_{\ell=1}^k X_{(i-1)m+\ell}$ .

### 2.2.1 Nonoverlapping Batch Means Estimator

One can show that as  $m \rightarrow \infty$ , the nonoverlapping batch means  $\bar{X}_{i,m}$  for  $i = 1, \dots, b$  become uncorrelated (Law and Carson 1979) and normally distributed when  $X$  obeys an FCLT. Then the NBM estimator for  $\sigma^2$  is defined by  $\mathcal{N}(b, m) \equiv m(b-1)^{-1} \sum_{i=1}^b (\bar{X}_{i,m} - \bar{X}_n)^2$ . If  $b$  is fixed, then  $\mathcal{N}(b, m) \xrightarrow[m \rightarrow \infty]{\text{d}} \sigma^2 \chi_{b-1}^2/(b-1)$ , and the usual NBM CI estimator for the mean,  $\bar{X}_n \pm t_{b-1, 1-\alpha/2} \sqrt{\mathcal{N}(b, m)/n}$ , is asymptotically valid as  $m \rightarrow \infty$  (Steiger and Wilson 2001). Under some additional assumptions, one can also show that

$$E[\mathcal{N}(b, m)] = \sigma^2 - \frac{(b+1)\gamma_1}{bm} + o\left(\frac{1}{m}\right) \quad \text{and} \quad \text{Var}[\mathcal{N}(b, m)] = \frac{2\sigma^2(b+1)}{(b-1)^2} + O\left(\frac{1}{bm^{1/4}}\right) + o\left(\frac{1}{n}\right);$$

see Alexopoulos and Goldsman (2004). If both  $m$  and  $b$  tend to  $\infty$ , then the last two equations imply weak consistency for  $\mathcal{N}(b, m)$  because  $\text{MSE}[\mathcal{N}(b, m)] \rightarrow 0$ ; but they do not guarantee the validity of the CI for the mean,  $\bar{X}_n \pm t_{b-1, 1-\alpha/2} \sqrt{\mathcal{N}(b, m)/n}$ .

Methods based on sequences of batch sizes and batch counts that aim at computing consistent estimators for  $\sigma^2$  and asymptotically valid CIs for  $\mu$  have been developed during the last twenty years (Damerdjii 1994, Fishman and Yarberry 1997). The latter method takes  $O(n)$  time and  $O(\log_2 n)$  space, and has been implemented in the LABATCH.2 package (available at the Web site [www.or.unc.edu/~gfish/labatch.2.html](http://www.or.unc.edu/~gfish/labatch.2.html)). The ASAP3 method (Steiger et al. 2005) has a fixed upper limit on  $b$  but allows  $m \rightarrow \infty$  so as to deliver approximately valid CIs for  $\mu$  that satisfy user-specified requirements on absolute or relative precision. This package is available at the Web site [www.ise.ncsu.edu/jwilson](http://www.ise.ncsu.edu/jwilson). Skart (Tafazzoli and Wilson 2011) exploits separate adjustments to the classical batch-means CI to account for the effects on the distribution of the underlying Student's  $t$ -statistic arising from skewness and autocorrelation of the batch means. A Visual Basic implementation of Skart is available online via [www.ise.ncsu.edu/jwilson](http://www.ise.ncsu.edu/jwilson). Both ASAP3 and Skart also deliver a point estimator for the steady-state mean that is approximately free of initialization bias.

### 2.2.2 Standardized Time Series

The STS for the sample  $\{X_1, \dots, X_n\}$  is defined by

$$T_n(t) \equiv \frac{\lfloor nt \rfloor (\bar{X}_n - \bar{X}_{\lfloor nt \rfloor})}{\sigma \sqrt{n}} \quad \text{for } t \in [0, 1] \text{ and } n = 1, 2, \dots$$

Under Assumption FCLT, it can be shown that  $(\sqrt{n}(\bar{X}_n - \mu), \sigma T_n) \xrightarrow[n \rightarrow \infty]{\Rightarrow} (\sigma \mathcal{W}(1), \sigma \mathcal{B})$ , where  $\mathcal{B}(\cdot)$  is a standard Brownian bridge process that is independent of  $\mathcal{W}(1)$ , so that  $T_n(\cdot)$  and  $\bar{X}_n$  are asymptotically independent (Schruben 1983).

### 2.2.3 Batched Area Estimators

The STS from batch  $i$  is  $T_{i,m}(t) \equiv \lfloor mt \rfloor (\bar{X}_{i,m} - \bar{X}_{i,\lfloor mt \rfloor}) / (\sigma \sqrt{m})$  for  $i = 1, \dots, b$  and  $t \in [0, 1]$ . Under the same mild conditions as above, one has

$$\begin{aligned} & (\sqrt{m}(\bar{X}_{1,m} - \mu), \sqrt{m}(\bar{X}_{2,m} - \mu), \dots, \sqrt{m}(\bar{X}_{b,m} - \mu); \sigma T_{1,m}, \sigma T_{2,m}, \dots, \sigma T_{b,m}) \\ & \xrightarrow[m \rightarrow \infty]{\Rightarrow} (\sigma \xi_1, \sigma \xi_2, \dots, \sigma \xi_b; \sigma \mathcal{B}_0, \sigma \mathcal{B}_1, \dots, \sigma \mathcal{B}_{b-1}), \end{aligned} \quad (1)$$

where the  $\xi_i$  are i.i.d. standard normal random variables, and  $\mathcal{B}_s(\cdot)$  denotes a standard Brownian bridge defined in terms of  $\mathcal{W}(\cdot)$  on  $[s, s+1]$ , for  $s \in [0, b-1]$ . That is,

$$\mathcal{B}_s(t) \equiv \mathcal{W}(s+t) - \mathcal{W}(s) - t[\mathcal{W}(s+1) - \mathcal{W}(s)], \quad t \in [0, 1] \text{ and } s \in [0, b-1].$$

One can easily show that the Brownian bridges  $\mathcal{B}_0(\cdot), \mathcal{B}_1(\cdot), \dots, \mathcal{B}_{b-1}(\cdot)$  are independent.

Define the signed areas

$$Z_i(f; m) \equiv \frac{1}{m} \sum_{\ell=1}^m f(\ell/m) \sigma T_{i,m}(\ell/m) = \frac{1}{m^{3/2}} \sum_{\ell=1}^m f(\ell/m) \ell (\bar{X}_{i,m} - \bar{X}_{i,\ell}) \quad \text{for } i = 1, \dots, b.$$

Equation (1) ensures that

$$[Z_1(f; m), Z_2(f; m), \dots, Z_b(f; m)] \xrightarrow[m \rightarrow \infty]{\Rightarrow} \sigma [\xi_1, \xi_2, \dots, \xi_b] \quad (2)$$

so that the  $\{Z_i(f; m) : i = 1, \dots, b\}$  become i.i.d.  $\sigma \text{Nor}(0, 1)$  as  $m \rightarrow \infty$  (Alexopoulos et al. 2007b, Goldsman et al. 1990).

The *batched area* estimator for  $\sigma^2$  is defined by

$$\mathcal{A}(f; b, m) \equiv \frac{1}{b} \sum_{i=1}^b Z_i^2(f; m) \xrightarrow[m \rightarrow \infty]{\Rightarrow} \sigma^2 \chi_b^2 / b.$$

Under Assumptions A, one can show that for fixed  $b$ ,  $E[\mathcal{A}(f; b, m)] = \sigma^2 - [(F - \bar{F})^2 + \bar{F}^2] \gamma_1 / (2m) + o(1/m)$ ; and, assuming uniform integrability for the family  $\{\mathcal{A}^2(f; b, m) : m = 1, 2, \dots\}$ , we have  $\lim_{m \rightarrow \infty} \text{Var}[\mathcal{A}(f; b, m)] = 2\sigma^4/b$  (Foley and Goldsman 1999, Goldsman et al. 1990). If one chooses weights having  $F = \bar{F} = 0$ , the resulting area estimator is first-order unbiased since its bias is  $o(1/m)$ . Damerджи and Goldsman (1995) gave conditions such that the estimator  $\mathcal{A}(f; b, m)$  is strongly consistent as both  $b, m \rightarrow \infty$  in an appropriate fashion.

Popular choices of weight functions are

$$\left. \begin{aligned} f_0(t) &= \sqrt{12} \quad (\text{Schruben 1983}), \\ f_2(t) &= \sqrt{840}(3t^2 - 3t + 1/2) \quad (\text{Goldsman et al. 1990}), \text{ and} \\ f_{\cos, j}(t) &= \sqrt{8\pi j} \cos(2\pi jt) \quad j = 1, 2, \dots \quad (\text{Foley and Goldsman 1999}) \end{aligned} \right\} \quad \text{for } t \in [0, 1].$$

Schruben’s estimator has  $E[\mathcal{A}(f_0; b, m)] = \sigma^2 - 3\gamma_1/m + o(1/m)$ , while the remaining estimators are first-order unbiased. Foley and Goldsman (1999) show that the area estimators arising from the weight functions  $\{f_{\cos,j}(\cdot) : j = 1, 2, \dots\}$  are also asymptotically independent. Thus we can average the first  $k$  of these area estimators to obtain a first-order unbiased estimator for  $\sigma^2$  with more degrees of freedom:

$$\overline{\mathcal{A}}_k(f_{\cos}; b, m) \equiv k^{-1} \sum_{j=1}^k \mathcal{A}(f_{\cos,j}; b, m) \xrightarrow{m \rightarrow \infty} \sigma^2 \chi_{bk}^2 / (bk) \quad \text{for } k = 1, 2, \dots$$

We note that the second-order bias of the area estimators based on the  $f_{\cos,j}(\cdot)$  weight function increases as  $j$  increases; this is the subject of ongoing research, but does not appreciably affect our results for small  $j$ .

### 2.2.4 Overlapping Batch Means

In various NBM procedures, the von Neumann (1941) randomness and the Shapiro and Wilk (1965) normality tests are used to determine a batch size that is sufficiently large to ensure that the resulting batch means are approximately i.i.d. normal with mean  $\mu$ . Unfortunately in large-scale applications, we often find that the resulting batch means exhibit nonnegligible correlation and skewness (Steiger et al. 2005, Tafazzoli and Wilson 2011). Such correlation usually causes systematic underestimation of  $\text{Var}[\overline{X}_n]$  (that is, negative bias); and this effect coupled with nonnegligible skewness of the batch means can cause the final CI coverage to fall substantially below its nominal level. In ASAP3 and Skart, we sought to estimate this correlation and compensate for it by modeling the batch means as a stationary first-order autoregressive process; moreover in Skart we also sought to estimate the skewness of the batch means and then to compensate for it by applying a suitable adjustment to the final CI. This approach proved to effective in a broad diversity of applications.

In the development of SPSTS, we observed similar problems with the von Neumann randomness test and the Shapiro–Wilk normality test when those tests were applied to the signed areas  $\{Z_i(f; m)\}$  so as to determine a batch size  $m$  that is sufficiently large to ensure the  $\{Z_i(f; m)\}$  are approximately i.i.d. normal with variance  $\sigma^2$ . In particular, Aktaran-Kalaycı et al. (2007) find that the resulting STS area estimator  $\mathcal{A}(f_{\cos,j}; b, m)$  of the variance parameter  $\sigma^2$  can exhibit significant negative second-order bias; this effect sometimes negates the lack of first-order bias and causes the final CI coverage to fall below its nominal level. To address this issue, we also used variance estimators based on overlapping batch means, which often exhibit reasonable small-sample bias.

Given a “shift”  $d \in \{1, \dots, m-1\}$ , we define the overlapping batches  $\{X_{(i-1)d+1}, \dots, X_{(i-1)d+m}\}$  for  $i = 1, \dots, k$ , where  $k \equiv \lfloor (n-m)/d \rfloor + 1$ . Within this context we define  $b \equiv n/m$ , with the understanding that it does not denote the number of batches. The respective overlapping batch means are

$$\overline{X}_{i,m;d}^o \equiv \frac{1}{m} \sum_{j=1}^m X_{(i-1)d+j} \quad \text{for } i = 1, \dots, k,$$

and the OBM estimator for  $\sigma^2$  is

$$\mathcal{O}(b, m; d) \equiv \frac{m}{k-1} \sum_{i=1}^k (\overline{X}_{i,m;d}^o - \overline{X}_n)^2. \tag{3}$$

The assignment  $d = 1$  yields the estimator of Meketon and Schmeiser (1984). Under the assumptions for the NBM estimator, it can be shown that the OBM variance estimator has about the same mean as NBM, but  $\text{Var}[\mathcal{O}(b, m; 1)] \approx \frac{2}{3} \text{Var}[\mathcal{N}(b, m)]$ . Welch (1987) noted that the NBM and OBM estimators are related to classical spectral estimators and showed that partial overlapping (i.e., taking  $d = m/q$  for a relatively small positive integer  $q$ ) not only saves computational effort, but also retains most of the benefits of full overlapping (i.e., taking  $d = 1$ ). For example,  $\text{Var}[\mathcal{O}(b, m; m/4)] \approx 0.69 \text{Var}[\mathcal{N}(b, m)]$  and  $\text{Var}[\mathcal{O}(b, m; m/2)] \approx 0.75 \text{Var}[\mathcal{N}(b, m)]$ .

### 3 SEQUENTIAL PROCEDURE

SPSTS is an automated sequential procedure for estimating the mean  $\mu$  of a stationary process  $X$  satisfying Assumptions A. It starts with  $b = 40$  nonoverlapping batches of size  $m = 2048$  so that the initial simulation-generated time series  $\{X_i\}$  must have length not less than  $n = bm = 81,920$ . To determine a subsequent batch size  $m$  that is sufficiently large so that the  $\{Z_i(f; m)\}$  are approximately i.i.d.  $\sigma\text{Nor}(0, 1)$  (that is, an adequate degree of convergence to the asymptotic result (2) has been achieved), SPSTS uses sequential versions of the von Neumann (1941) randomness test and the Shapiro and Wilk (1965) normality test. If the  $\{Z_i(f; m)\}$  fail iteration  $r$  of the von Neumann randomness test ( $r = 1, 2, \dots$ ) at the significance level (type I error)  $\beta\psi(r)$ , then the batch size  $m$  is increased by a factor of about  $\sqrt{2}$ , the iteration counter  $r$  is incremented, and the randomness test is reperformed. We use  $\beta = 0.20$  and  $\psi(r) \equiv \exp[-\eta(r-1)^\theta]$ ; and the current values of the parameters  $\eta$  and  $\theta$  are  $\eta = 0.184206$  and  $\theta = 2$  as specified in Lada and Wilson (2006).

Step [5] below addresses the simulation *start-up problem* by truncating the first batch of observations. The idea behind this truncation is that once the signed areas pass the randomness test in steps [2]–[4], one can assume that the observations beyond the first batch are free of initialization bias. Steps [6]–[8] apply the Shapiro–Wilk normality test. If the  $\{Z_i(f; m)\}$  fail iteration  $\ell$  of the normality test ( $\ell = 1, 2, \dots$ ) at the significance level  $\delta\psi(\ell)$ , with  $\delta = 0.20$ , then the batch size  $m$  is increased by the factor  $\sqrt{2}$ , the iteration counter  $\ell$  is incremented, and the normality test is reperformed.

Given the user-specified relative precision  $\gamma > 0$ , SPSTS sequentially increases the total sample size  $n$ , the batch size  $m$ , and the batch count  $b$  (up to the limit of  $b^* = 64$  batches) so that in terms of the final point estimator  $\bar{X}_n$  of  $\mu$ , the half-length of the final CI for  $\mu$  does not exceed  $\gamma|\bar{X}_n|$ . When  $\gamma = \infty$  (no precision case), the algorithm terminates as soon as the signed areas based on the cosine weights  $f_{\cos,1}(\cdot)$  and  $f_{\cos,2}(\cdot)$  pass both tests for randomness and normality. Finite values of  $\gamma$  should be reasonably small, say  $\gamma \leq 0.15$ . A formal algorithmic statement of SPSTS is given below.

#### Algorithm SPSTS

- 
- [0] *Initialization:* Set  $\alpha$ ,  $\beta = 0.20$ ,  $\delta = 0.20$ ,  $b^* = 64$ , and relative error upper bound  $\gamma$ . Define the function  $\psi(r) \equiv \exp[-\eta(r-1)^\theta]$ ,  $r = 1, 2, \dots$ , where  $\eta > 0$  and  $\theta \geq 1$ .
  - [1] Generate  $b = 40$  batches of size  $m = 2048$ . Set  $j \leftarrow 1$  and  $r \leftarrow 1$ .
  - [2] **Until** von Neumann’s test fails to reject independence using the current weight function  $f_j(\cdot)$ :
    - compute the signed areas  $\{Z_i(f_j; m) : i = 1, \dots, b\}$  for weight function  $f_j(t) \equiv f_{\cos,j}(t)$ ,  $t \in [0, 1]$ .
    - assess the independence of  $\{Z_i(f_j; m) : i = 1, \dots, b\}$  using von Neumann’s two-sided randomness test at significance level  $\beta\psi(r)$ ;
    - set  $r \leftarrow r + 1$  and  $m \leftarrow \lfloor m\sqrt{2} + 0.5 \rfloor$ ;
  - [3] Set  $m_j \leftarrow m$ ;  $m \leftarrow 2048$ ;  $j \leftarrow 2$ ; and  $r \leftarrow 1$ .
  - [4] Repeat step [2] with the updated value of  $j$ ; and take  $m \leftarrow \max\{m_1, m_2\}$ .
  - [5] Remove the first batch  $\{X_1, \dots, X_m\}$ ; reindex the remaining sample and the signed areas as  $\{Z_i(f_j; m) : i = 1, \dots, b-1; j = 1, 2\}$ ; collect a new batch of size  $m$ ; compute  $\{Z_b(f_j; m) : j = 1, 2\}$ ; and set  $j \leftarrow 1$ .
  - [6] **Until** the Shapiro–Wilk test fails to reject normality using the current weight function  $f_j(\cdot)$ :
    - assess the multivariate normality of  $\{Z_i(f_j; m)\}$  using the Shapiro–Wilk one-sided test at significance level  $\delta\psi(\ell)$ ;
    - set  $\ell \leftarrow \ell + 1$  and  $m \leftarrow \lfloor m\sqrt{2} + 0.5 \rfloor$ ;

- [7] Set  $m'_j \leftarrow m$ ;  $m \leftarrow \max\{m_1, m_2\}$ ;  $j \leftarrow 2$ ; and  $\ell \leftarrow 1$ .  
 [8] Repeat step [6] with the updated value of  $j$  and take  $m \leftarrow \max\{m'_1, m'_2\}$ .  
 [9] **Until** the relative half-length  $h(b, m, \alpha)/|\bar{X}_n| \leq \gamma$ :

- compute the batched area estimates

$$\mathcal{A}(f_j; b, m) \equiv \frac{1}{b} \sum_{i=1}^b Z_i^2(f_j; m) \quad \text{for } j = 1, 2;$$

- compute the OBM variance estimate  $\mathcal{O}(b, m; m/4)$  using Equation (3) with shift  $d = m/4$ ;
- set

$$\widehat{V}(b, m) \leftarrow \max \{ \mathcal{O}(b, m; m/4), \mathcal{A}(f_j; b, m) : j = 1, 2 \}$$

and compute the half-length

$$h(b, m, \alpha) \equiv t_{b, 1-\alpha/2} \sqrt{\widehat{V}(b, m)/n};$$

- estimate the number of batches of the current size required to satisfy the precision requirement

$$b' = \left\lceil \left\{ \frac{h(b, m, \alpha)}{\gamma \bar{X}_n} \right\}^2 \right\rceil;$$

- update the batch count  $b$ , the batch size  $m$ , and the total sample size  $n$  as follows,

$$\begin{aligned} b &\leftarrow \min \{ b', b^* \}, \\ m &\leftarrow \begin{cases} m, & \text{if } b = b', \\ \lceil m \times \text{mid}\{1.05, (b'/b), 2\} \rceil, & \text{if } b < b', \end{cases} \\ n &\leftarrow bm. \end{aligned}$$

- [10] Deliver the  $100(1 - \alpha)\%$  CI  $\bar{X}_n \pm h(b, m, \alpha)$ .

---

**Remarks:** The factor of  $\sqrt{2}$  applied to the increase of the batch size in steps [2], [4], [6], and [8] allows computational savings because accumulators collected in iteration  $r$  are used to compute the  $\{Z_i(f_j; m)\}$  in iteration  $r+2$ . Further, for each  $i$  and  $m$ , the quantities  $Z_i(f_1; m)$  and  $Z_i(f_2; m)$  are computed in parallel.

#### 4 COMPARISONS WITH EXISTING METHODS

This section discusses a preliminary empirical study designed to test the performance of SPSTS. To do so, we consider a stationary  $M/M/1$  queueing system with arrival rate  $\omega$  and traffic intensity  $\rho < 1$ . Let  $X_i$  be the time spent in queue prior to service by entity  $i$ . For this process, we have  $\mu = \rho^2/[\omega(1 - \rho)]$  and  $\sigma^2 = \rho^3(2 + 5\rho - 4\rho^2 + \rho^3)/[\omega^2(1 - \rho)^4]$ ; see, for example, Steiger and Wilson (2001). Table 1 contains experimental results for  $\omega = 0.9$  and  $\rho = 1$ . The estimates for ASAP3 and Skart are from Table 1 of Tafazzoli et al. (2011). All three methods used the initial condition  $X_0 = 0$ . The results for SPSTS and Skart are based on 1000 replications of each procedure, whereas the results for ASAP3 are based on 400 replications of that procedure.

Table 1 reports results on the following performance characteristics for the SPSTS, Skart, and ASAP3 procedures for CIs with no precision requirement and precisions of 7.5% and 3.75%: empirical coverage,

average sample size until procedure termination, average CI half-length, and sample standard deviation of CI half-length. For the specific  $M/M/1$  process under study, the CI coverage probabilities for SPSTS always met or (slightly) exceeded the user-specified nominal levels. Under the no-precision requirement, this was achieved at the cost of substantially larger sample sizes than Skart and ASAP3. (The reader should bear in mind that SPSTS starts with a significantly larger sample size than Skart or ASAP3.) On the other hand, the CIs delivered by SPSTS were less variable than their competitors from Skart and ASAP3. Under the rather stringent precision requirements of 7.5% and 3.75%, the sample sizes required by SPSTS were about the same as those required for Skart, but somewhat higher than the corresponding sample sizes for ASAP3. Furthermore, the CIs delivered by SPSTS were noticeably more variable than their counterparts from ASAP3 and Skart for the relative precision of 3.75%. This variability can be addressed by area estimators based on STSs applied to overlapping batches (Alexopoulos et al. 2007b).

Table 1: Experimental results for the SPSTS algorithm for an  $M/M/1$  system with traffic intensity 0.9. The estimates for SPSTS and Skart are based on 1000 independent replications, while the estimates for ASAP3 are based on 400 replications.

Precision	Performance Measure	Nominal 90% CIs			Nominal 95% CIs		
		SPSTS	Skart	ASAP3	SPSTS	Skart	ASAP3
none ( $\gamma = \infty$ )	CI coverage (%)	92.5	87.6	87.5	95.6	93.9	91.5
	Avg. sample size	291,790	42,369	31,181	291,790	42,369	31,181
	Avg. CI half-length	0.638	1.767	2.072	0.765	2.298	2.521
	StDev. CI half-length	0.153	0.508	0.590	1.184	1.137	0.731
$\pm 7.5\%$	CI coverage (%)	92.1	91.1	89.5	95.9	95.9	94.0
	Avg. sample size	326,440	302,305	287,568	417,700	431,677	382,958
	Avg. CI half-length	0.576	0.635	0.627	0.604	0.637	0.632
	StDev. CI half-length	0.082	0.037	0.048	0.069	0.035	0.045
$\pm 3.75\%$	CI coverage (%)	91.4	92.0	89.5	95.9	96.0	93.5
	Avg. sample size	1,146,900	1,105,417	969,011	1,588,500	1,586,267	1,341,522
	Avg. CI half-length	0.305	0.321	0.320	0.310	0.321	0.321
	StDev. CI half-length	0.027	0.014	0.020	0.022	0.014	0.020

## 5 CONCLUSIONS

This paper introduced SPSTS, an automated sequential procedure for computing point and CI estimators for the steady-state mean of a simulation output process. SPSTS is based on STS area estimators; and in a small empirical evaluation involving an  $M/M/1$  queueing system, we found that SPSTS performed reasonably well compared with other sequential procedures in terms of CI coverage, average required sample size, and the first two moments of the CI's half-length. It should be noted that STS area estimators can also be used for obtaining CIs for the variance parameter  $\sigma^2$  (Alexopoulos et al. 2007a). The following tasks will be undertaken in the near future: (a) we will develop sequential procedures based on STS area estimators computed from overlapping batches of observations (Alexopoulos et al. 2007b); (b) we will enhance our procedures to address the start-up problem for simulations containing a transient phase; and (c) in order to evaluate the robustness and competitiveness of the new methodology more completely, we plan on carrying out a great deal of additional Monte Carlo experimentation on a large variety of stochastic processes with more-challenging characteristics — including processes with heavy tails, long-term memory, etc.

## ACKNOWLEDGMENTS

Support for this work has been provided by National Science Foundation grants CMMI-1232998/1233141 and EFRI-ARES-CI-0735991.

## REFERENCES

- Aktaran-Kalaycı, T., C. Alexopoulos, N. T. Argon, D. Goldsman, and J. R. Wilson. 2007. "Exact Expected Values of Variance Estimators in Steady-State Simulation". *Naval Research Logistics* 54 (4): 397–410.
- Alexopoulos, C., N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. 2007a. "Efficient Computation of Overlapping Variance Estimators for Simulation". *INFORMS Journal on Computing* 19 (3): 314–327.
- Alexopoulos, C., N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. 2007b. "Overlapping Variance Estimators for Simulation". *Operations Research* 55 (6): 1090–1103.
- Alexopoulos, C., and D. Goldsman. 2004. "To Batch or Not to Batch?". *ACM Transactions on Modeling and Computer Simulation* 14 (1): 76–114.
- Antonini, C., C. Alexopoulos, D. Goldsman, and J. R. Wilson. 2009. "Area Variance Estimators for Simulation Using Folded Standardized Time Series". *IIE Transactions* 41:134–144.
- Damerджи, H. 1994. "Strong Consistency of the Variance Estimator in Steady-State Simulation Output Analysis". *Mathematics of Operations Research* 19:494–512.
- Damerджи, H., and D. Goldsman. 1995. "Consistency of Several Variants of the Standardized Time Series Area Variance Estimator". *Naval Research Logistics* 42:1161–1176.
- Durrett, R. 2005. *Probability: Theory and Examples*. 3rd ed. Belmont, CA: Brooks/Cole–Thomson Learning.
- Fishman, G. S. 2001. *Discrete-Event Simulation: Modeling, Programming, and Analysis*. New York: Springer-Verlag.
- Fishman, G. S., and L. S. Yarbber. 1997. "An Implementation of the Batch Means Method". *INFORMS Journal on Computing* 9:296–310.
- Foley, R. D., and D. Goldsman. 1999. "Confidence Intervals Using Orthonormally Weighted Standardized Time Series". *ACM Transactions on Modeling and Simulation* 9:297–325.
- Fuller, W. A. 1996. *Introduction to Statistical Time Series*. 2nd ed. New York: John Wiley & Sons.
- Glynn, P. W., and D. L. Iglehart. 1990. "Simulation Output Analysis Using Standardized Time Series". *Mathematics of Operations Research* 15:1–16.
- Goldsman, D., M. S. Meketon, and L. W. Schruben. 1990. "Properties of Standardized Time Series Weighted Area Variance Estimators". *Management Science* 36:602–612.
- Lada, E. K., and J. R. Wilson. 2006. "A Wavelet-Based Spectral Procedure for Steady-State Simulation Analysis". *European Journal of Operational Research* 174 (3): 1769–1801.
- Law, A. M. 2007. *Simulation Modeling and Analysis*. 4th ed. New York: McGraw-Hill.
- 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*, edited by S. Sheppard, U. W. Pooch, and C. D. Pegden, 227–230. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Schruben, L. W. 1983. "Confidence Interval Estimation Using Standardized Time Series". *Operations Research* 31:1090–1108.
- Shapiro, S. S., and M. B. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)". *Biometrika* 52:591–611.
- Song, W.-M., and B. W. Schmeiser. 1995. "Optimal Mean-Squared-Error Batch Sizes". *Management Science* 41:110–123.
- Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2005. "ASAP3: A Batch Means Procedure for Steady-State Simulation Analysis". *ACM Transactions on Modeling and Computer Simulation* 15 (1): 39–73.
- 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.
- Tafazzoli, A., and J. R. Wilson. 2011. "Skart: A Skewness- and Autoregression-Adjusted Batch Means Procedure for Simulation Analysis". *IIE Transactions* 43 (2): 110–128.

- von Neumann, J. 1941. "Distribution of the Ratio of the Mean Square Successive Difference to the Variance". *Annals of Mathematical Statistics* 12 (4):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*, edited by A. Thesen, H. Grant, and W. D. Kelton, 320–323. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

#### AUTHOR BIOGRAPHIES

**CHRISTOS ALEXOPOULOS** is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests are in the areas of simulation, statistics, and optimization of stochastic systems. He is a member of INFORMS and an active participant in the Winter Simulation Conference, having been proceedings co-editor in 1995, associate program chair in 2006, and a member of the Board of Directors since 2008. He is also an area editor of the *ACM Transactions on Modeling and Computer Simulation*. His e-mail address is [christos@isye.gatech.edu](mailto:christos@isye.gatech.edu), and his Web page is [www.isye.gatech.edu/~christos](http://www.isye.gatech.edu/~christos).

**DAVID GOLDSMAN** is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests include simulation output analysis, ranking and selection, and healthcare simulation. He was program chair of the Winter Simulation Conference in 1995 and a member of the WSC Board of Directors between 2001–2009. He is currently a trustee of the WSC Foundation. His e-mail address is [sman@gatech.edu](mailto:sman@gatech.edu), and his Web page is [www.isye.gatech.edu/~sman](http://www.isye.gatech.edu/~sman).

**PENG TANG** is a doctoral student in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. Her interests are in applied statistics and simulation output analysis. Her email address is [ptang7@isye.gatech.edu](mailto:ptang7@isye.gatech.edu).

**JAMES R. WILSON** is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. His current research interests are focused on probabilistic and statistical issues in the design and analysis of simulation experiments, with special emphasis on applications in healthcare and production. As a WSC participant, he served as proceedings editor (1986), associate program chair (1991), and program chair (1992). During the period 1997–2004, he was a member of the WSC Board of Directors. He is a member of ACM, ASA, and SCS; and he is a Fellow of IIE and INFORMS. His e-mail address is [jwilson@ncsu.edu](mailto:jwilson@ncsu.edu), and his Web page is [www.ise.ncsu.edu/jwilson](http://www.ise.ncsu.edu/jwilson).