

## STATISTICAL ANALYSIS OF SIMULATION OUTPUT: STATE OF THE ART

Christos Alexopoulos

H. Milton Stewart School of Industrial and Systems Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332, U.S.A.

### ABSTRACT

This paper reviews statistical methods for analyzing output data from computer simulations. Specifically, it focuses on the estimation of steady-state system parameters. The estimation techniques include the replication/deletion approach, the regenerative method, the batch means method, and methods based on standardized time series.

### 1 INTRODUCTION

A primary goal of most simulation studies is the approximation of prescribed system parameters with the objective of identifying parameter values that optimize some system performance measures. Since the input processes driving a simulation are often random, the output data are also random and runs of the simulation program only result in *estimates* of system performance measures.

A simulation study consists of several steps such as data collection, coding and verification, model validation, experimental design, output data analysis, and implementation. This paper reviews statistical methods for computing confidence intervals for system performance measures from output data. This tutorial does not aim at replacing “standard” texts, such as Fishman (2001) or Law (2006). A comprehensive coverage of the majority of topics reviewed herein is presented in Chapters 8, 15, and 16 of the recent handbook edited by Henderson and Nelson (2006) and in the entry by Alexopoulos et al. (2006b).

There are two types of simulations with regard to output analysis:

**Finite-horizon simulations.** In this case the simulation starts in a specific state and is run until some terminating event occurs. The output process is not expected to achieve any steady-state behavior and any parameter estimated from the output data will be transient in the sense that its value will depend upon the initial conditions. An example is the simulation of a vehicle storage and distribution facility for a week.

**Steady-state simulations.** The purpose of a steady-state simulation is the study of the long-run behavior of the system of interest. A performance measure of a system is called a *steady-state parameter* if it is a characteristic of the equilibrium distribution of an output stochastic process. An example is the simulation of a continuously operating communication system where the objective is the computation of the mean delay of a data packet.

Section 2 discusses methods for analyzing output from finite-horizon simulations. Section 3 presents techniques for point and interval estimation of steady-state parameters.

### 2 FINITE-HORIZON SIMULATIONS

Suppose that we simulate a system until  $n$  output data  $X_1, X_2, \dots, X_n$  are collected with the objective of estimating  $\mu \equiv E(\bar{X}_n)$ , where  $\bar{X}_n \equiv \frac{1}{n} \sum_{i=1}^n X_i$  is the sample mean of the data. For example,  $X_i$  may be the transit time of unit  $i$  through a network of queues or the total time station  $i$  is busy during the  $i$ th hour. Clearly,  $\bar{X}_n$  is an unbiased estimator for  $\mu$ . Unfortunately, the  $X_i$  are generally dependent random variables making the estimation of the variance  $\text{Var}(\bar{X}_n)$  a nontrivial problem. Let  $S_n^2(X) \equiv \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  be the sample variance of the data. The presence of autocorrelation makes the familiar estimator  $S_n^2(X)/n$  a biased estimator of  $\text{Var}(\bar{X}_n)$  (see Section 3).

To overcome this problem, one can run  $k$  independent replications of the system simulation. Assume that run  $i$  produces the output data  $X_{i1}, X_{i2}, \dots, X_{in}$ . Then the replicate averages  $Y_i \equiv \frac{1}{n} \sum_{j=1}^n X_{ij}$  are independent and identically distributed (IID) random variables, their sample mean  $\bar{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i$  is also an unbiased estimator of  $\mu$ , and their sample variance  $S_k^2(Y)$  is an unbiased estimator of  $\text{Var}(\bar{X}_n)$ . If in addition  $k$  is sufficiently large, an approximate  $100(1 - \alpha)\%$  two-sided confidence interval (CI) for  $\mu$  is

$$\bar{Y}_k \pm t_{k-1, 1-\alpha/2} S_k(Y) / \sqrt{k}, \quad (1)$$

where  $t_{d,\delta}$  represents the  $\delta$ -quantile of Student's  $t$  distribution with  $d$  degrees of freedom.

Alexopoulos and Seila (1998, Section 7.2.2) review sequential procedures for determining the number of replications required to estimate  $\mu$  with a fixed absolute or relative precision. The procedure for computing an estimate  $\bar{Y}_k$  with an absolute error satisfying  $\Pr(|\bar{Y}_k - \mu| \leq \beta) \geq 1 - \alpha$  is based on Chow and Robbins (1965). It starts with  $k \geq 5$  runs and stops when the halfwidth  $t_{k-1,1-\alpha/2}S_k(Y)/\sqrt{k} \leq \beta$ . Law (2006, pp. 501–502) describes an empirical method for obtaining an estimate whose relative error satisfies  $\Pr(|\bar{Y}_k - \mu|/|\mu| \leq \gamma) \geq 1 - \alpha$ , with  $\alpha \leq 0.15$ . The method starts with  $k \geq 10$  runs and stops when the relative halfwidth  $t_{k-1,1-\alpha/2}S_k(Y)/(|\bar{Y}_k|\sqrt{k})$  drops below  $\gamma/(1+\gamma)$ .

The method of replications can also be used for estimating performance measures other than means. Let  $Y$  be the total cost incurred in an inventory system during a certain time window, and let  $y_p \equiv \inf\{y : \Pr(Y \leq y) \geq p\}$  denote the  $p$ -quantile of  $Y$ . To estimate  $y_p$ , we can make  $k$  independent replications, denote by  $Y_i$  the cost observed during replication  $i$ , and let  $Y_{(1)} < Y_{(2)} < \dots < Y_{(k)}$  be the order statistics corresponding to the  $Y_i$ . Then a point estimate for  $y_p$  is  $\hat{y}_p = Y_{(\lceil kp \rceil)}$ , where  $\lceil \cdot \rceil$  is the ceiling function. A CI for  $y_p$  is described in Alexopoulos and Seila (1998, Section 7.3.2).

### 3 STEADY-STATE ANALYSIS

We focus on methods for computing point and interval estimators for the mean of a discrete-time stationary process. Analogous methods for analyzing continuous-time output data are described in a variety of texts (Fishman 2001; Law 2006). The process  $X \equiv \{X_i\}$  is called *stationary* if the joint distribution of  $X_{i+j_1}, X_{i+j_2}, \dots, X_{i+j_k}$  is independent of  $i$  for all indices  $j_1, j_2, \dots, j_k$  and all  $k \geq 1$ . If  $E(X_i) = \mu$ ,  $\text{Var}(X_i) \equiv \sigma_X^2 < \infty$  for all  $i$ , and the  $\text{Cov}(X_i, X_{i+j})$  is independent of  $i$ , then  $X$  is called *weakly stationary*. We denote the autocovariance function of  $X$  by  $R_j \equiv \text{Cov}(X_1, X_{1+j})$  ( $j = 0, \pm 1, \pm 2, \dots$ ). Notice that  $R_0 = \sigma_X^2$ . Alexopoulos et al. (2006b) give a detailed overview of the properties of stationary processes that are central to simulation output analysis and proceed with a detailed description of the majority of the methods in this section. Clearly, the sample mean  $\bar{X}_n$  is not only unbiased for  $\mu$ , but also strongly consistent by the ergodic theorem (see Durrett 2005).

Under the assumption that  $\bar{X}_n$  is approximately normally distributed (which is reasonable for sufficiently large  $n$ ), the usual construction of a CI for  $\mu$  requires the derivation of an estimator for  $\text{Var}(\bar{X}_n)$ . A little algebra yields (cf. Anderson 1984)

$$E \left[ \frac{S_n^2(X)}{n} \right] = \frac{n-1}{n-1} \text{Var}(\bar{X}_n), \quad (2)$$

where  $a_n \equiv 1 + (2/\sigma_X^2) \sum_{j=1}^{n-1} (1-j/n)R_j$ . Then for processes that are positively correlated ( $R_j > 0$ ), equation (2) implies that  $E[S_n^2(X)/n] < \text{Var}(\bar{X}_n)$ . Hence the “classical”  $100(1-\alpha)\%$  CI for IID data  $\bar{X}_n \pm t_{n-1,1-\alpha/2}S_n(X)/\sqrt{n}$  can have coverage probability that can be considerably below the nominal value  $1-\alpha$ .

A common assumption facilitating the derivation of a CI for  $\mu$  is as follows:

**Functional Central Limit Theorem (FCLT) Assumption.** Suppose that the series

$$\sigma^2 \equiv \sigma_X^2 + 2 \sum_{j=1}^{\infty} R_j \quad (3)$$

is absolutely convergent and  $\sigma^2 > 0$ . Let

$$X_n(t) \equiv \frac{\lfloor nt \rfloor (\bar{X}_{\lfloor nt \rfloor} - \mu)}{\sigma \sqrt{n}}, \quad t \in [0, 1],$$

where  $\lfloor \cdot \rfloor$  is the floor function. Then  $X_n(\cdot) \xrightarrow{\mathcal{D}} \mathcal{W}$ , where “ $\xrightarrow{\mathcal{D}}$ ” denotes convergence in distribution and  $\{\mathcal{W}(t) : 0 \leq t \leq 1\}$  is a standard Brownian motion process restricted to the interval  $[0, 1]$ . We call  $\sigma^2$  the (*asymptotic*) *variance parameter* of  $X$ .

This assumption holds under several conditions (Durrett 2005; Glynn and Iglehart 1990). Examples are a condition involving conditional second moments of  $X$  and the stronger  $\varphi$ -mixing condition:  $X$  is  $\varphi$ -mixing if there are  $\varphi_k \downarrow 0$  such that, for each  $k \geq 0$ ,  $A \in \mathcal{F}_{-\infty}^j$ , and  $B \in \mathcal{F}_{j+k}^{\infty}$ ,  $|\Pr(A \cap B) - \Pr(A)\Pr(B)| \leq \varphi_k \Pr(A)$ . Here  $\mathcal{F}_i^j$  ( $i \leq j$ ) denotes the  $\sigma$ -field generated by  $X_i, X_{i+1}, \dots, X_j$ .

**Remark 1** Contrary to popular belief, many stochastic processes encountered in simulation output analysis are *not*  $\varphi$ -mixing. Examples are autoregressive processes, regenerative processes (see Section 3.3) with regenerations not occurring uniformly fast over the state space, and virtually all open queueing networks (Glynn and Iglehart 1985).

The variance of the sample mean in terms of the autocovariance function is

$$\text{Var}(\bar{X}_n) = \frac{1}{n} \left[ \sigma_X^2 + 2 \sum_{j=1}^{n-1} (1-j/n)R_j \right]. \quad (4)$$

Assumption  $0 < \sigma^2 < \infty$  along with equation (4) imply  $\lim_{n \rightarrow \infty} n \text{Var}(\bar{X}_n) = \sigma^2$  and  $\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = 0$ ; hence  $\bar{X}_n$  is consistent (in mean square). Our focus will be on methods for obtaining CIs for  $\mu$ , which involve estimating  $\sigma^2$ .

Below the “little-oh” notation  $f(m) = o(g(m))$  means that  $f(m)/g(m) \rightarrow 0$  as  $m \rightarrow \infty$ ; and the “big-oh” notation

$f(m) = O(g(m))$  means that there is a positive integer  $m_0$  such that  $|f(m)/g(m)| \leq C$  for some constant  $C$  and all  $m \geq m_0$ .

### 3.1 Dealing with the Initial Conditions

Several problems arise when the process  $X$  does not start in steady-state. For example,  $\bar{X}_n$  is not an unbiased estimator of the mean  $\mu$ . The removal of the effect of the initial conditions is a challenging problem.

The most commonly used method for eliminating the bias of  $\bar{X}_n$  identifies an index  $l$  and “truncates” the observations  $X_1, \dots, X_l$ . Several procedures have been proposed for the detection of a cutoff index  $l$  (see Fishman 2001; Law 2006; Wilson and Pritsker 1978ab).

The graphical procedure of Welch (1983) uses  $k$  independent replications, with the  $i$ th replication producing observations  $X_{i1}, X_{i2}, \dots, X_{in}$ , and computes the “across-runs” averages  $\tilde{X}_j \equiv \frac{1}{k} \sum_{i=1}^k X_{ij}$  ( $j = 1, \dots, n$ ). Then for a given time window  $w$ , the procedure plots the *moving averages*

$$\bar{X}_j(w) = \begin{cases} \frac{1}{2w+1} \sum_{m=-w}^w \tilde{X}_{j+m} & w+1 \leq j \leq n-w \\ \frac{1}{2j-1} \sum_{m=-j+1}^{j-1} \tilde{X}_{j+m} & 1 \leq j \leq w \end{cases}$$

against  $j$ . If the plot is reasonably smooth, then  $l$  is chosen to be the value of  $j$  beyond which the sequence of moving averages converges. Otherwise, a different time window is chosen and a new plot is drawn. The choice of  $w$  may be a difficult problem for congested systems with output time series having autocorrelation functions with long tails (see Alexopoulos and Seila 1998, Example 7).

### 3.2 The Replication/Deletion Approach

This intuitive, and often recommended, approach runs  $k$  independent replications, each of length  $l+n$  observations, and discards the first  $l$  observations from each run. One then uses the IID sample means  $Y_i(l, n) \equiv \frac{1}{n} \sum_{j=l+1}^{l+n} X_{ij}$  from the  $k$  runs to compute the point estimate  $\bar{Y}_k(l, n) \equiv \frac{1}{k} \sum_{i=1}^k Y_i(l, n)$  and the following approximate  $100(1-\alpha)\%$  CI for  $\mu$ :

$$\bar{Y}_k(l, n) \pm t_{k-1, 1-\alpha/2} S_k(Y(l, n)) / \sqrt{k}, \quad (5)$$

where  $S_k^2(Y(l, n))$  is the sample variance of the  $Y_i(l, n)$ .

The method is simple and general, but involves the choice of three parameters,  $l$ ,  $n$  and  $k$ . Here are a few points the user should be aware of: (a) As  $l$  increases for fixed  $n$ , the “systematic” error in each  $Y_i(l, n)$  due to the initial conditions decreases. (b) As  $n$  increases for fixed  $l$ , the systematic and sampling errors in  $Y_i(l, n)$  decrease. (c) The systematic error in the sample means  $Y_i(l, n)$  cannot

be reduced by increasing the number of replications  $k$ . (d) For fixed  $n$  and under some mild moment conditions that are satisfied by a variety of simulation output processes, a sufficient condition for the asymptotic validity of the CI (5) is  $l / \ln k \rightarrow \infty$  as  $k \rightarrow \infty$  (Fishman 2001). This means that as one makes more runs in an attempt to compute a narrower CI, the truncation index  $l$  should increase faster than  $\ln k$  for the CI to achieve the nominal coverage. This requirement is hard to implement in practice. (e) This method is also potentially wasteful of data as the truncated portion is removed from each replication. Additional shortcomings of this method are presented in Alexopoulos and Goldsman (2004).

The methods in Sections 3.3–3.6 seek to overcome the aforementioned issues. For a thorough comparison between the methods of independent replications and nonoverlapping batch means, see Alexopoulos and Goldsman (2004).

### 3.3 The Regenerative Method

This method assumes the identification of time indices at which the process  $X$  probabilistically *starts over* and uses these regeneration epochs for obtaining IID random variables which can be used for computing point and interval estimates for the mean  $\mu$ . The method was proposed by Crane and Iglehart (1975) and Fishman (1973, 1974). More precisely, assume that there are (random) time indices  $1 \leq T_1 < T_2 < \dots$  such that the portion  $\{X_{T_i+j} : j \geq 0\}$  has the same distribution for each  $i$  and is independent of the portion prior to time  $T_i$ . The portion of the process between two successive regeneration epochs is called a *cycle*. Let  $Y_i = \sum_{j=T_i}^{T_{i+1}-1} X_j$  and  $Z_i = T_{i+1} - T_i$  ( $i = 1, 2, \dots$ ) and assume that  $E(Z_i) < \infty$ . Then the steady-state mean  $\mu$  is given by  $\mu = E(Y_1)/E(Z_1)$ .

Now suppose that one simulates the process  $X$  over  $n$  cycles and collects the observations  $Y_1, \dots, Y_n$  and  $Z_1, \dots, Z_n$ . Then  $\hat{\mu} \equiv \bar{Y}_n / \bar{Z}_n$  is a strongly consistent estimator of  $\mu$ . Furthermore, CIs for  $\mu$  can be constructed by using the IID random variables  $Y_i - \mu Z_i$ , ( $i = 1, \dots, n$ ) and the central limit theorem (see Iglehart 1975).

The regenerative method is often difficult to apply in practice because the majority of simulations have either no regenerative points or very long cycle lengths. Two classes of systems this method has successfully been applied to are inventory systems and highly reliable communications systems with repairs.

### 3.4 Methods Based on Nonoverlapping Batch Means

The classical method of nonoverlapping batch means (NBM) is a popular approach for computing point and CI estimators for the mean  $\mu$  of a stationary process. Original accounts on the method were given by Conway (1963), Fishman (1978), and Law and Carson (1979); see Fishman (2001) and Alexopoulos and Goldsman (2004) for detailed coverage.

Suppose that the sample  $\{X_1, \dots, X_n\}$  is divided into  $k$  contiguous batches, each consisting of  $m$  observations (for simplicity, we assume  $n = km$ ). For  $i = 1, \dots, k$ , the  $i$ th batch consists of the observations  $X_{(i-1)m+1}, X_{(i-1)m+2}, \dots, X_{im}$  and the  $i$ th batch mean  $Y_{i,m} \equiv \frac{1}{m} \sum_{j=1}^m X_{(i-1)m+j}$  is the sample average from batch  $i$ . The NBM-based estimator of the mean is the grand sample mean  $\bar{X}_n = \frac{1}{k} \sum_{i=1}^k Y_{i,m} = \frac{1}{n} \sum_{i=1}^n X_i$ . Clearly, the stationarity of  $X$  implies  $E(\bar{X}_n) = \mu$  and the stationarity of the batch means sequence  $\{Y_{i,m} : i = 1, \dots, k\}$ .

The motivation behind the NBM method is simple. First, under the FCLT, one can show that as  $m \rightarrow \infty$ , the batch means become uncorrelated (Law and Carson 1979) and normally distributed. Since the grand mean  $\bar{X}_n$  is the sample average of the batch means, one has the approximation  $n\text{Var}(\bar{X}_n) \doteq n\text{Var}(Y_{1,m})/k = m\text{Var}(Y_{1,m})$ . Hence the NBM estimator for  $\sigma^2$  is

$$\mathcal{N}(k, m) \equiv \frac{m}{k-1} \sum_{i=1}^k (Y_{i,m} - \bar{X}_n)^2, \quad (6)$$

which is  $m$  times the sample variance of the batch means. An approximate  $100(1 - \alpha)\%$  CI for  $\mu$  is

$$\bar{X}_n \pm t_{k-1, 1-\alpha/2} \sqrt{\frac{\mathcal{N}(k, m)}{n}}. \quad (7)$$

Of course, the fundamental issue is the choice of the batch size and the number of batches. Several early studies (e.g., Fishman 1978; Schmeiser 1982) addressed this issue, but without the rigor of recent studies.

To motivate the description of the modern procedures, we focus on the mean squared error (MSE) of  $\mathcal{N}(k, m)$  and the coverage of the CI (7). Here we let  $\sigma_n^2 \equiv n\text{Var}(\bar{X}_n)$ , and define the constant

$$\gamma \equiv - \sum_{j=-\infty}^{\infty} j R_j = -2 \sum_{j=1}^{\infty} j R_j. \quad (8)$$

One can show easily that  $E[\mathcal{N}(k, m)] = \frac{1}{k-1}(k\sigma_m^2 - \sigma_n^2)$ . If in addition  $E(X_1^4) < \infty$ , and the process  $X$  is  $\varphi$ -mixing with  $\varphi_j = O(j^{-4-\epsilon})$  for some  $\epsilon > 0$ , then  $\gamma$  exists and

$$\sigma_n^2 = \sigma^2 + \gamma/n + o(1/n). \quad (9)$$

Combining the last two equations we obtain

$$E[\mathcal{N}(k, m)] = \sigma^2 + (k+1)\gamma/n + o(1/n). \quad (10)$$

Hence,  $\mathcal{N}(k, m)$  usually has negative first-order bias for positively autocorrelated processes.

Also, the additional assumptions  $E(X_1^2) < \infty$  and  $\varphi_j = O(j^{-9})$  allow one to write (Chien et al. 1997)

$$\text{Var}[\mathcal{N}(k, m)] = \frac{2\sigma^4(k+1)}{(k-1)^2} + O\left(\frac{1}{km^{1/4}}\right) + O\left(\frac{1}{k^2}\right). \quad (11)$$

Then, as  $m, k \rightarrow \infty$ ,

$$\text{MSE}[\mathcal{N}(k, m)] = O\left(\frac{1}{km^{1/4}}\right) + O\left(\frac{1}{k^2}\right) \rightarrow 0. \quad (12)$$

Property (12) implies weak consistency for the estimator  $\mathcal{N}(k, m)$ , but does not guarantee the asymptotic validity of the CI in equation (7). Before we discuss batching rules that yield the last property, we briefly examine how the variance estimator  $\mathcal{N}(k, m)$  approaches  $\sigma^2$ . As in Fishman (2001, p. 251), equation (10) allows us to write

$$\begin{aligned} \mathcal{N}(k, m) - \sigma^2 &= \underbrace{\sigma_n^2 - \sigma^2}_{\text{error due to finite } n} - \underbrace{\sigma_n^2 \frac{1 - \sigma_m^2/\sigma_n^2}{1 - m/n}}_{\text{error due to ignoring correlations between batch means}} \\ &+ \underbrace{\epsilon_n}_{\text{error due to random sampling}}, \end{aligned} \quad (13)$$

where the error  $\epsilon_n$  has mean zero and variance given by equation (11). We call the first two terms on the right-hand side of the last equation a systematic error; by equation (9) this error behaves as  $O(1/m)$ . On the other hand, equation (11) implies that the standard deviation of  $\epsilon_n$  behaves as  $O(1/k^{1/2})$ . These growth rates reveal the tradeoff between the two types of error induced by  $m$  and  $k$ . Since  $\sigma_n^2$  approaches  $\sigma^2$  from below for a variety of systems with positive autocorrelation functions, the systematic error induces a negative bias in  $\mathcal{N}(k, m)$  that dissipates as the batch size increases. Then the error due to random sampling fluctuates around zero and decreases at rate  $O(1/k^{1/2})$ .

The recent literature contains a variety of rules for selecting sequences of batch sizes  $\{m_\ell\}$  and batch counts  $\{k_\ell\}$  as the sample size increases. The most intuitive (FNB) rule fixes the number of batches and doubles the batch size at each iteration. This assignment is computationally attractive because at every iteration, pairs of existing batch means are averaged to compute the new batch means.

Under the FCLT assumption, one can show that for fixed  $k$  and  $m \rightarrow \infty$ ,  $\mathcal{N}(k, m) \xrightarrow{\mathcal{D}} \sigma^2 \chi_{k-1}^2 / (k-1)$ , where  $\chi_d^2$  denotes a chi-square random variable with  $d$  degrees of freedom; and the CI in equation (7) is asymptotically valid (Glynn and Whitt 1991). If we assume uniform integrability for the sequence  $\{\mathcal{N}^2(k, m) : m \geq 1\}$  (see Billingsley 1968), we have  $\lim_{m \rightarrow \infty} E[\mathcal{N}(k, m)] = \sigma^2$  and

$\lim_{m \rightarrow \infty} \text{Var}[\mathcal{N}(k, m)] = 2\sigma^4/(k-1)$ ; hence the FNB rule does not yield a consistent variance estimator. This is in agreement with equation (13) as the error  $O(k^{-1/2})$  due to random sampling does not diminish. Therefore the CI in equation (7) tends to be wider than CIs based on consistent variance estimators.

### Consistent Batch Means Estimation Methods

Alternative rules that yield strongly consistent estimators for  $\mathcal{N}(k, m)$  are based on the following assumption:

**Assumption of Strong Approximation (ASA).** There exists a constant  $\lambda \in (0, 1/2]$  and a finite random variable  $C$  such that, as  $n \rightarrow \infty$ ,  $|\sqrt{n}(\bar{X}_n - \mu)/\sigma - \mathcal{W}(n)/\sqrt{n}| \leq Cn^{-\lambda}$ , w.p.1, where  $\mathcal{W}$  is a standard Brownian motion process defined on the same space as the standardized process  $\{\bar{X}_n\}$ .

A  $\lambda$  close to  $1/2$  indicates a marginal normal distribution and low correlation among the  $X_i$ . Conversely, a  $\lambda$  close to zero indicates the absence of at least one of these properties (Philipp and Stout 1975). The following theorem proposes batching assumptions which along with ASA yield a strongly consistent estimator for  $\sigma^2$ . (Notice that the batching sequences are indexed by the sample size.)

**Theorem 1** (Damerdj 1994a) *Suppose that the ASA holds and that  $\{m_n\}$  and  $\{k_n\}$  are deterministic sequences of batch sizes and batch counts, respectively, such that  $m_n \rightarrow \infty$ ,  $k_n \rightarrow \infty$ ,  $n^{1-2\lambda} \ln(n)/m_n \rightarrow 0$  (as  $n \rightarrow \infty$ ), and  $\sum_{n=1}^{\infty} k_n^{-q} < \infty$  for some finite integer  $q \geq 1$ . Then, as  $n \rightarrow \infty$ ,  $\mathcal{N}(k_n, m_n) \rightarrow \sigma^2$ , w.p.1 and*

$$Z(k_n, m_n) \equiv \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mathcal{N}(k_n, m_n)}} \xrightarrow{\mathcal{D}} N(0, 1), \quad (14)$$

where  $N(0, 1)$  is a standard normal random variable.

Suppose that  $m_n \doteq n^\theta$ , for some  $\theta \in (0, 1)$ . One can verify that the conditions of Theorem 1 are satisfied if  $\theta \in (1 - 2\lambda, 1)$ . In particular, the square root (SQRT) rule that uses  $m_n \doteq k_n \doteq \sqrt{n}$  ( $\theta = 1/2$ ) yields a strongly consistent variance estimator when  $1/4 < \lambda < 1/2$ . In addition to the derivation of a strongly consistent estimator for  $\sigma^2$ , the SQRT rule induces an optimal property: Assuming that  $E(X_1^{20}) < \infty$  and that  $X$  is  $\varphi$ -mixing with  $\varphi_j = O(j^{-13})$ , Chien (1989) showed that the CDF of the standardized statistic  $Z(k, m)$  converges to the standard normal CDF at the fastest possible rate. Unfortunately, the CIs for  $\mu$  that result from an implementation of the SQRT rule often exhibit low coverage for small sample sizes (see Example 11 in Alexopoulos and Seila 1998).

Although both the FNB and SQRT rules yield asymptotically valid CIs for  $\mu$ , each has desirable properties and limitations. To close the gap, Fishman and Yarbber (1997) proposed the LABATCH.2 suite of algorithms. Among the two recommended algorithms, LBATCH and ABATCH, we present the latter because it is more conservative with re-

gard to the coverage of the resulting CI (7). This method uses von Neumann's test (von Neumann 1941) to assess the hypothesis  $H_0$ : "the batch means are IID." The associated test statistic is

$$\Gamma(k, m) \equiv \sqrt{\frac{k^2 - 1}{k - 2}} \left[ 1 - \frac{\sum_{i=1}^{k-1} (Y_{i,m} - Y_{i+1,m})^2}{2 \sum_{i=1}^k (Y_{i,m} - \bar{X}_n)^2} \right].$$

Under  $H_0$ ,  $\Gamma(k, m) \sim N(0, 1)$  for large  $m$  (the batch means become approximately normal) or large  $k$ . The symbol " $\sim$ " means "is approximately distributed as." To guard against positive correlation, one can use a one-sided test and reject  $H_0$  at level  $\beta$  when  $\Gamma(k, m) > z_{1-\beta}$ , where  $z_\delta$  is the  $\delta$ -quantile of the standard normal distribution.

The ABATCH algorithm evolves as follows. For a complete description, see Fishman (2001).

### Algorithm ABATCH

- Select initial batch size  $m_1$ , initial batch count  $k_1$ , confidence level  $1 - \alpha$ , and type I error  $\beta$  for von Neumann's test.
- On iteration  $\ell \geq 1$ : Compute von Neumann's statistic  $\Gamma(k_\ell, m_\ell)$ . If  $\Gamma(k_\ell, m_\ell) > z_{1-\beta}$ , reject  $H_0$  and use the FNB rule on iteration  $\ell + 1$  with  $k_{\ell+1} = k_\ell$  and  $m_{\ell+1} = 2m_\ell$ . Otherwise, use the SQRT rule on iteration  $\ell + 1$  with  $k_{\ell+1} \doteq \sqrt{2}k_\ell$  and  $m_{\ell+1} \doteq \sqrt{2}m_\ell$ .

Since the ABATCH algorithm uses random  $m_\ell$ 's and  $k_\ell$ 's, Theorem 6.6 of Fishman (2001) lists additional mild conditions that imply strong consistency for  $\mathcal{N}(k_\ell, m_\ell)$  and asymptotic validity for the CI  $\bar{X}_n \pm z_{1-\alpha/2} \sqrt{\mathcal{N}(k_\ell, m_\ell)/n_\ell}$  as  $\ell \rightarrow \infty$ . The FNB and SQRT rules can be implemented easily within the ABATCH algorithm by setting  $\beta = 0$  or  $\beta = 1$ , respectively. Two features of the LABATCH.2 suite that are often overlooked are algorithm efficiency and low space requirements: each algorithm requires  $O(n)$  total time and  $O(\log_2 n)$  space. Although like complexities are known for static fixed-batch-size algorithms (e.g., all the methods in the remainder of this paper have a linear time complexity per iteration), the dynamic setting of ABATCH 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, in linear total time. This assessment is essential to gauge the quality of the variance parameter estimates and the CI for the mean and requires a rather experienced user. C, FORTRAN and SIMSCRIPT II.5 codes of LABATCH.2 can be downloaded via anonymous ftp from the site [www.or.unc.edu/~gfish/labatch.2.html](http://www.or.unc.edu/~gfish/labatch.2.html).

The ASAP3 method of Steiger et al. (2005) is an automated sequential NBM approach that delivers a CI for  $\mu$  satisfying user-specified requirements on absolute or relative precision as well as coverage probability. This approach

takes advantage of the fact that the batch means often become approximately multivariate normal random variables before achieving independence. ASAP3 operates as follows. The batch size is progressively 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 inverted Cornish-Fisher expansion for the classical batch means  $t$ -ratio based on the AR(1) parameter estimates; and finally ASAP3 delivers a correlation-adjusted CI based on this expansion. Although ASAP3 does not possess the computational efficiency of the LABATCH.2 algorithms, it performs very well with regard to conformance to the precision and coverage probability requirements as well as with regard to the mean and variance of the half-length of the delivered CI. Related papers, experimental results, and the ASAP3 software are accessible from the site [www.ise.ncsu.edu/jwilson](http://www.ise.ncsu.edu/jwilson). A detailed experimental study of sequential NBM procedures is presented in Lada et al. (2006).

A reasonable compromise between the methods of independent replications (IR) in Section 2 and NBM has been proposed recently by Argon and Andradóttir (2006). The *replicated batch means* (RBM) method uses a few independent replications of equal length, each containing the same number of batches, and estimates the variance parameter  $\sigma^2$   $m$  times the sample variance of all batch means. When the output process is stationary, the RBM method appears to exhibit performance characteristics that fall between the constituent IR and NBM methods. The recent paper by Alexopoulos et al. (2006a) studies the performance of the RBM variance estimator in the presence of an additive transient bias.

### 3.5 Estimators Based on Standardized Time Series

Now we turn to estimators based on standardized time series (STS). We start with estimators based on the entire sample, and then present estimators based on standardized time series applied to batches. The STS based on the sample  $\{X_1, \dots, X_n\}$  is defined as

$$T_n(t) \equiv \frac{\lfloor nt \rfloor (\bar{X}_n - \bar{X}_{\lfloor nt \rfloor})}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1.$$

If  $X$  satisfies an FCLT, it can be shown that, as  $n \rightarrow \infty$ ,

$$(\sqrt{n}(\bar{X}_n - \mu), \sigma T_n) \xrightarrow{\mathcal{D}} (\sigma \mathcal{W}(1), \sigma \mathcal{B}), \quad (15)$$

where  $\mathcal{B}$  is the standard Brownian bridge process on  $[0, 1]$  defined by  $\mathcal{B}(t) = \mathcal{W}(t) - t\mathcal{W}(1)$ . In addition, the STS  $T_n(\cdot)$  is asymptotically independent of  $\bar{X}_n$ . All finite-dimensional

joint distributions of  $\mathcal{B}$  are normal with  $E[\mathcal{B}(t)] = 0$  and  $\text{Cov}[\mathcal{B}(s), \mathcal{B}(t)] = \min(s, t) - st$ ,  $0 \leq s, t \leq 1$ .

### The Weighted Area Estimator

We start with the *weighted area estimator* (Goldsman et al. 1990; Goldsman and Schruben 1990; Schruben 1983). Suppose that the function  $f$  is twice continuously differentiable on the interval  $[0, 1]$  and normalized so that  $\text{Var}(\int_0^1 f(t)\mathcal{B}(t) dt) = 1$ . Then  $\int_0^1 f(t)\mathcal{B}(t) \sim \sigma N(0, 1)$ . The square of the weighted area under the STS is defined by

$$A(f; n) \equiv \left[ \frac{1}{n} \sum_{i=1}^n f(i/n) \sigma T_n(i/n) \right]^2.$$

Under mild conditions, the continuous mapping theorem (see Billingsley 1968, Theorem 5.1) implies

$$A(f; n) \xrightarrow{\mathcal{D}} A(f) \equiv \left[ \int_0^1 f(t) \sigma \mathcal{B}(t) dt \right]^2 \sim \sigma^2 \chi_1^2,$$

as  $n \rightarrow \infty$ . The following theorem gives expressions for the mean and variance of the weighted area estimator  $A(f; n)$ . Notice that the limiting variance does not depend on the weight function  $f$ .

**Theorem 2** (Goldsman et al. 1990; Foley and Goldsman 1999) *Suppose that  $X$  is  $\varphi$ -mixing and satisfies Assumption FCLT,  $\sum_{j=1}^{\infty} j^2 |R_j| < \infty$ , and the sequence  $\{A^2(f; n) : n \geq 1\}$  is uniformly integrable. Then, as  $n \rightarrow \infty$ ,*

$$E[A(f; n)] = \sigma^2 + \frac{[(F(1) - \bar{F}(1))^2 + \bar{F}^2(1)]\gamma}{2n} + o(1/n)$$

and

$$\text{Var}[A(f; n)] \rightarrow \text{Var}[A(f)] = \text{Var}(\sigma^2 \chi_1^2) = 2\sigma^4,$$

where  $F(s) \equiv \int_0^s f(t) dt$  ( $0 \leq s \leq 1$ ) and  $\bar{F}(u) \equiv \int_0^u F(s) ds$  ( $0 \leq u \leq 1$ ).

**Example 1** Schruben (1983) studied the area estimator with constant weight function  $f_0(t) \equiv \sqrt{12}$ ; in this case, Theorem 2 implies that  $E[A(f_0; n)] = \sigma^2 + 3\gamma/n + o(1/n)$ .

If one chooses weights having  $F(1) = \bar{F}(1) = 0$ , the resulting estimator is *first-order unbiased* for  $\sigma^2$ , i.e., its bias is  $o(1/n)$ . An example of a weight function yielding a first-order unbiased estimator for  $\sigma^2$  is  $f_2(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$  (Goldsman et al. 1990; Goldsman and Schruben 1990).

Other weights yielding first-order unbiased estimators for  $\sigma^2$  are given by the family  $f_{\cos, j}(t) = \sqrt{8\pi j} \cos(2\pi jt)$  ( $j = 1, 2, \dots$ ). Foley and Goldsman (1999) showed that this *orthonormal* sequence produces area estima-

tors  $A(f_{\cos,1}, n), A(f_{\cos,2}, n), \dots$  that are not only first-order unbiased, but asymptotically independent; that is,  $A(f_{\cos,1}), A(f_{\cos,2}), \dots$  are IID  $\sigma^2 \chi_1^2$ .

### Batched Area Estimators

Up to now, the STS-based variance estimators have been constructed directly from a single long run of  $n$  observations. We now examine what happens if we (a) divide the run into contiguous, nonoverlapping *batches*; (b) form an STS estimator from each batch; and (c) take the average of the estimators.

The STS from batch  $i$  ( $i = 1, \dots, k$ ) is

$$T_{i,m}(t) \equiv \frac{\lfloor mt \rfloor (Y_{i,m} - Y_{i,\lfloor mt \rfloor})}{\sigma \sqrt{m}}, \quad 0 \leq t \leq 1,$$

where  $Y_{i,j} \equiv \frac{1}{j} \sum_{\ell=1}^j X_{(i-1)m+\ell}$ . Under the same mild conditions as before, one has

$$\begin{aligned} & (\sqrt{m}(Y_{1,m} - \mu), \sqrt{m}(Y_{2,m} - \mu), \dots, \sqrt{m}(Y_{k,m} - \mu); \\ & \sigma T_{1,m}, \sigma T_{2,m}, \dots, \sigma T_{k,m}) \\ & \xrightarrow{\mathcal{D}} (\sigma Z_1, \sigma Z_2, \dots, \sigma Z_k; \sigma \mathcal{B}_0, \sigma \mathcal{B}_1, \dots, \sigma \mathcal{B}_{k-1}), \end{aligned}$$

where the  $Z_i$  are IID standard normal random variables and

$$\mathcal{B}_s(t) \equiv \mathcal{W}(s+t) - \mathcal{W}(s) - t[\mathcal{W}(s+1) - \mathcal{W}(s)], \quad 0 \leq t \leq 1$$

denotes a standard Brownian bridge on  $[s, s+1]$ , for  $s \in [0, k-1]$ . One can easily show that the Brownian bridges  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$  are independent.

The area estimator from batch  $i$  is

$$A_i(f; m) \equiv \left[ \frac{1}{m} \sum_{\ell=1}^m f(\ell/m) \sigma T_{i,m}(\ell/m) \right]^2, \quad i = 1, \dots, k,$$

and the (nonoverlapping) *batched area* estimator for  $\sigma^2$  is

$$\mathcal{A}(f; k, m) \equiv \frac{1}{k} \sum_{i=1}^k A_i(f; m). \quad (16)$$

Since the  $T_{i,m}$  ( $i = 1, \dots, k$ ), converge to independent Brownian bridges as  $m$  becomes large (with  $k$  fixed), the  $A_i(f; m)$  are asymptotically independent as  $m \rightarrow \infty$ . Then by the discussion above, we have  $\mathcal{A}(f; k, m) \xrightarrow{\mathcal{D}} \sigma^2 \chi_k^2 / k$ , and an approximate  $100(1 - \alpha)\%$  CI for  $\mu$  is  $\bar{X}_n \pm t_{k-1, 1-\alpha/2} \sqrt{\mathcal{A}(f; k, m) / n}$ .

Theorem 2 implies

$$\begin{aligned} E[\mathcal{A}(f; k, m)] &= \sigma^2 + \frac{[(F(1) - \bar{F}(1))^2 + \bar{F}^2(1)]\gamma}{2m} \\ &+ o(1/m). \end{aligned} \quad (17)$$

Further, if we assume uniform integrability for  $\{\mathcal{A}^2(f; k, m) : m \geq 1\}$ , we can also make an analogous statement concerning the variance of the batched area estimator: as  $m \rightarrow \infty$ ,

$$\begin{aligned} \text{Var}[\mathcal{A}(f; k, m)] &= k^{-1} \text{Var}[A_1(f; m)] \\ &\rightarrow k^{-1} \text{Var}[A(f)] = 2\sigma^4 / k. \end{aligned} \quad (18)$$

Equations (17) and (18) indicate that the batched area estimator has a bit more bias than the area estimator obtained from the entire sample, but smaller asymptotic variance (by a factor of  $k$ ). Sargent et al. (1992) present an extensive experimental study for various CIs mentioned in this section.

It is worth mentioning that, under the assumptions of Theorem 1, Damerdjji (1994ab,1995) showed that the batched area estimator  $\mathcal{A}(f; k, m)$  is strongly and mean-square consistent. Methods based on NBMs and batched area estimators have been used for computing point and CI estimators for continuous nonlinear functions of steady-state means (Muñoz and Glynn 1997; Chang 2004).

### 3.6 Methods Based on Overlapping Batches

Now we turn our attention to methods relying on overlapping batches, starting with the method of *overlapping* batch means (OBM) proposed by Meketon and Schmeiser (1984). For given batch size  $m$ , this method uses  $n - b + 1$  overlapping batches to estimate  $\mu$  and  $\text{Var}(\bar{X}_n)$ . The first batch consists of observations  $X_1, \dots, X_m$ , the second batch consists of  $X_2, \dots, X_{m+1}$ , etc. The point estimator for  $\mu$  remains  $\bar{X}_n$ , but the OBM-based estimator of  $\sigma^2$  is

$$\mathcal{O}(k, m) = \frac{nm}{(n-m+1)(n-k)} \sum_{i=1}^{n-m+1} (Y_{i,m}^O - \bar{X}_n)^2,$$

where  $Y_{i,m}^O \equiv \frac{1}{m} \sum_{\ell=i}^{i+m-1} X_\ell$  ( $i = 1, \dots, n-m+1$ ) are the respective batch means and  $k \equiv n/m$  is no longer the number of batches. Actually, this estimator is almost identical to Bartlett's spectral estimator (see Anderson 1984).

Under conditions similar to those required to derive equations (10) and (11) one has (Song and Schmeiser 1995)

$$E[\mathcal{O}(k, m)] = \sigma^2 + \frac{(k^2 + 1)\gamma}{k(k-1)m} + o(1/m) \quad (19)$$

and

$$\text{Var}[\mathcal{O}(k, m)] \rightarrow \frac{(4k^3 - 11k^2 + 4k + 6)\sigma^4}{3(k-1)^4} \doteq \frac{4\sigma^4}{3k}, \quad (20)$$

as  $m \rightarrow \infty$  (Damerdjji 1995, Appendix B). Equations (10) and (19) show that the estimators  $\mathcal{N}(k, m)$  and  $\mathcal{O}(k, m)$  have the same asymptotic means (as  $k, m \rightarrow \infty$ ). How-

ever, a comparison between equations (11) and (20) reveals that  $\text{Var}[\mathcal{O}(k, m)]/\text{Var}[\mathcal{N}(k, m)] \rightarrow 2/3$ , as  $k, m \rightarrow \infty$ . Thus, the OBM method has better (asymptotic) performance than NBM with regard to MSE. Also, the behavior of  $\text{Var}[\mathcal{O}(k, m)]$  appears to be less sensitive to the choice of the batch size than does the behavior of  $\text{Var}[\mathcal{N}(k, m)]$  (see Song and Schmeiser 1995, Table 1).

An approximate  $100(1 - \alpha)\%$  CI for  $\mu$  is  $\bar{X}_n \pm t_{d, 1-\alpha/2} \sqrt{\mathcal{O}(k, m)/n}$ , with the degrees of freedom  $d$  chosen so that  $\mathcal{O}(k, m)$  is asymptotically  $\sigma^2 \chi_d^2/d$ . Meketon and Schmeiser (1984) use the value  $d = 1.5(k - 1)$  whereas, based on Monte Carlo studies, Schmeiser recommends the larger value  $d = 1.5(k - 1)[1 + (k - 1)^{-0.5-0.6k}]$ .

The OBM method can also yield a consistent variance estimator. If  $X$  satisfies ASA, the batching sequences satisfy the assumptions of Theorem 1, and  $\lim_{n \rightarrow \infty} (k_n^2/n) = 0$ , then  $\text{Var}[\mathcal{O}(k_n, m_n)] \rightarrow \sigma^2$ , w.p.1 (Damerджи 1994a).

Using equations (19) and (20), one can show that for a sample size  $n$ , the batch size that minimizes the  $\text{MSE}[\mathcal{O}(k, m)]$  is given by  $m^* = [3\gamma^2 n / (2\sigma^4)]^{1/3}$ . Song (1996) developed methods for estimating the ratio  $\gamma^2/\sigma^4$  for a variety of processes, including moving average processes and autoregressive processes. Then one can obtain an estimator for  $m^*$  by plugging the ratio estimator into the respective formula.

Welch (1987) suggested that overlapping batch means yield near-optimal variance reduction under partial overlapping. For example, under a 75% overlap with a batch size of 64, the first batch consists of observations  $X_1, \dots, X_{64}$ , the second consists of observations  $X_{17}, \dots, X_{80}$ , etc.

### Overlapping Area Estimators

To parallel the discussion in Section 3.5, we define the STS from overlapping batch  $i$  as

$$T_{i,m}^{\mathcal{O}}(t) \equiv \frac{[mt](Y_{i,m}^{\mathcal{O}} - Y_{i,[mt]}^{\mathcal{O}})}{\sigma \sqrt{m}}, \quad 0 \leq t \leq 1,$$

where  $Y_{i,j}^{\mathcal{O}} \equiv \frac{1}{j} \sum_{\ell=0}^{j-1} X_{i+\ell}$  ( $i = 1, \dots, n - m + 1$ ). The area estimator computed from overlapping batch  $i$  is defined by

$$A_i^{\mathcal{O}}(f; m) \equiv \left[ \frac{1}{m} \sum_{\ell=1}^m f(\ell/m) \sigma T_{i,m}^{\mathcal{O}}(\ell/m) \right]^2$$

and the *overlapping area* estimator for  $\sigma^2$  is

$$\mathcal{A}^{\mathcal{O}}(f; k, m) \equiv \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} A_i^{\mathcal{O}}(f; m). \quad (21)$$

The following discussion contains results from Alexopoulos et al. (2007ab).

**Theorem 3** Under Assumption FCLT, as  $m \rightarrow \infty$ ,

$$\mathcal{A}^{\mathcal{O}}(f; k, m) \xrightarrow{\mathcal{D}} \mathcal{A}^{\mathcal{O}}(f; k) \equiv \frac{1}{k-1} \int_0^{k-1} \left[ \sigma \int_0^1 f(u) \mathcal{B}_s(u) du \right]^2 ds.$$

Clearly, the expected value of the estimator  $\mathcal{A}^{\mathcal{O}}(f; k, m)$  matches that of its nonoverlapping counterpart. On the other hand, contrary to equation (18), the variance of the overlapping area estimator *does* depend on the choice of the weight function. For each weight function under consideration, assume that the family  $\{[\mathcal{A}^{\mathcal{O}}(f; k, m)]^2 : m \geq 1\}$  is uniformly integrable. For the constant weight function  $f_0$  in Example 1, one has

$$\begin{aligned} \text{Var}[\mathcal{A}^{\mathcal{O}}(f_0; k, m)] &\rightarrow \text{Var}[\mathcal{A}^{\mathcal{O}}(f_0; k)] \\ &= \frac{24k - 1}{35(k - 1)^2} \sigma^4, \quad \text{as } m \rightarrow \infty. \end{aligned}$$

>From the last result and equation (18) one obtains  $\text{Var}[\mathcal{A}^{\mathcal{O}}(f_0; k, m)]/\text{Var}[\mathcal{A}(f_0; k, m)] \doteq 0.343$  for large  $k$ . Alternatively, the first-order unbiased quadratic weight  $f_2$  yields

$$\text{Var}[\mathcal{A}^{\mathcal{O}}(f_2; k)] = \frac{3514k - 4359}{4290(k - 1)^2} \sigma^4,$$

hence  $\text{Var}[\mathcal{A}^{\mathcal{O}}(f_2; k, m)]/\text{Var}[\mathcal{A}(f_2; k, m)] \doteq 0.410$  for large  $k$ . Finally, for the orthonormal family of weights  $f_{\cos, j}$  in Example 1, one has

$$\text{Var}[\mathcal{A}^{\mathcal{O}}(f_{\cos, j}; k)] \doteq \frac{8\pi^2 j^2 + 15}{12\pi^2 j^2 k} \sigma^4.$$

Thus, for large  $k$ ,  $\text{Var}[\mathcal{A}^{\mathcal{O}}(f_{\cos, j}; k, m)]/\text{Var}[\mathcal{A}(f_{\cos, j}; k, m)] \doteq 0.397$  for  $j = 1$  and decreases to  $1/3$  as  $j \rightarrow \infty$ . Moreover, these variance estimators usually achieve similar improvements compared with the conventional estimators  $\mathcal{N}(k, m)$  and  $\mathcal{O}(k, m)$ .

In addition to the aforementioned results for the first two moments of the overlapping variance estimators, and under some additional assumptions, Alexopoulos et al. (2007ab) obtain consistency in mean square as  $m, k \rightarrow \infty$  and present efficient algorithms for computing these estimators with order-of-sample-size work. Aktaran-Kalaycı et al. (2006) present detailed expressions for the expected values of various variance estimators listed above.

Although Theorem 3 displays the asymptotic distribution of  $\mathcal{A}^{\mathcal{O}}(f; k, m)$ , one wonders if there are practical approximations to that distribution that can be used for constructing (among others) CIs for  $\mu$  and  $\sigma^2$ . Notice that for sufficiently large  $m$ ,  $\mathcal{A}^{\mathcal{O}}(f; k, m)$  is the average of approximate chi-squared random variables. Using the technique of Satterthwaite (1941), one can obtain the distribution



approximation

$$\mathcal{A}^O(f; k, m) \sim E[\mathcal{A}^O(f; k, m)]\chi_{\nu_{\text{eff}}}^2/\nu_{\text{eff}}$$

$$\text{where } \nu_{\text{eff}} = \left\lceil \frac{2E^2[\mathcal{A}^O(f; k, m)]}{\text{Var}[\mathcal{A}^O(f; k, m)]} \right\rceil, \quad (22)$$

where  $\lceil \cdot \rceil$  denotes rounding towards the nearest integer. The quantity  $\nu_{\text{eff}}$  is called the “effective” degrees of freedom.

If  $\mathcal{A}^O(f; k, m)$  is a first-order unbiased estimator for  $\sigma^2$ , then it follows from equation (22) that for a sufficiently large batch size  $m$ , an approximate  $100(1 - \alpha)\%$  CI for  $\sigma^2$  is given by

$$\frac{\nu_{\text{eff}} \mathcal{A}^O(f; k, m)}{\chi_{\nu_{\text{eff}}, 1-\alpha/2}^2} \leq \sigma^2 \leq \frac{\nu_{\text{eff}} \mathcal{A}^O(f; k, m)}{\chi_{\nu_{\text{eff}}, \alpha/2}^2}, \quad (23)$$

where  $\chi_{d, \delta}^2$  denotes the  $\delta$ -quantile of the  $\chi_d^2$  distribution.

We can also obtain an approximate CI for the mean  $\mu$ , provided the batch size  $m$  is sufficiently large (so that the sample size  $n$  is also sufficiently large). First of all, equation (15) implies that for large  $n$ ,  $\bar{X}_n \sim N(\mu, \sigma^2/n)$  and  $\bar{X}_n$  is approximately independent of the standardized time series formed from  $\{X_1, \dots, X_n\}$ . It follows that for  $m$  sufficiently large, the pivot

$$\frac{\bar{X}_n - \mu}{\sqrt{\mathcal{A}^O(f; k, m)/n}} \sim t_{\nu_{\text{eff}}}, \quad (24)$$

with the degrees of freedom  $\nu_{\text{eff}}$  given by (22). Then an approximate  $100(1 - \alpha)\%$  CI for  $\mu$  is given by

$$\mu \pm t_{\nu_{\text{eff}}, 1-\alpha/2} \sqrt{\mathcal{A}^O(f; k, m)/n}. \quad (25)$$

**Example 2** Consider the stationary Gaussian first-order autoregressive (AR(1)) process defined by  $X_i = 0.9X_{i-1} + \varepsilon_i$  for  $i = 1, 2, \dots$ , where  $X_0$  is a  $N(0, 1)$  random variable and the  $\varepsilon_i$ 's are IID  $N(0, 1 - \phi^2)$  random variables that are independent of  $X_0$ . This process has covariance function  $R_j = 0.9^j$  for all  $j = 0, 1, \dots$ , so that  $\sigma^2 = 19$ . The experimental results below come from Sections 4.1–4.3 of Alexopoulos et al. (2007a) and are based on 1,000,000 independent sample paths, each consisting of 20,000 observations, and a batch size of  $m = 1000$ .

For the first-order unbiased overlapping area estimator  $\mathcal{A}^O(f_2; 20, 1000)$ , one has  $E[\mathcal{A}^O(f_2; 20, 1000)] \approx \sigma^2 = 19$  and  $\nu_{\text{eff}} = 47$ . Figure 1 displays the empirical density, represented as a polygon as detailed in Section 3.3 of Hald (1952), and the fitted density based on (22). The excellent fit of the scaled chi-square approximation is apparent.

On the other hand, equation (23) yields the following approximate 90% CI for  $\sigma^2$ :

$$0.7343\mathcal{A}^O(f_2; 20, 1000) \leq \sigma^2 \leq 1.4566\mathcal{A}^O(f_2; 20, 1000).$$

The empirical coverage of this CI, based on the same 1,000,000 sample paths, was 0.9013, which is remarkably close to the nominal coverage.

Lastly, using  $\nu_{\text{eff}} = 47$  in equation (25), one has the following approximate 90% CI for  $\mu = 0$ :  $\bar{X}_n \pm 0.01187\sqrt{\mathcal{A}^O(f_2; 20, 1000)/n}$ . The empirical coverage of the last interval was 0.8992.

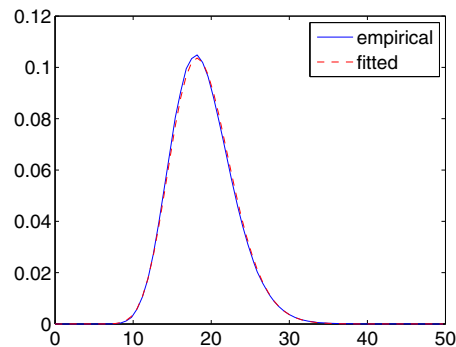


Figure 1: Empirical and Fitted Densities for the Overlapping Area Estimator  $\mathcal{A}^O(f_2; 20, 1000)$  Computed from an AR(1) Process with  $\phi = 0.9$

### Other Variance Estimation Methods

The list of estimators above is not exhaustive. The Cramér-von Mises estimators in Goldsman et al. (1999) and Alexopoulos et al. (2007ab) are based on the weighted area under the square of the STS. Antonini et al. (2007) propose estimators based on iterative “folding” operations applied to STSs. The purpose of the folding operation is the derivation of STSs that are asymptotically independent (as the sample size approaches infinity) despite the fact that they are based on the same output data. Hence, averaging of variance estimators computed from the various STSs yields an estimator with substantially smaller mean squared error than each of the constituents. Calvin and Nakayama (2006) propose estimators obtained by permuting STSs, and Calvin (2007) develops estimators based on integrated sample paths.

We end this section with a brief review of spectral methods. Classical spectral methods (cf. Heidelberg and Welch 1981, 1983) are based on the fact that the asymptotic variance parameter  $\sigma^2$  is equal to the value of the power spectrum of the stochastic process  $X$  at zero frequency. Hence, they work in the frequency domain to estimate  $\sigma^2$  from the values of the associated periodogram in the vicin-

ity of zero frequency. The WASSP method of Lada and Wilson (2006) is based on nonoverlapping batch means and wavelets. The method starts by selecting a data-truncation point and a batch size so that the truncated batch means are approximately a stationary Gaussian process but are not necessarily independent. Then it computes the discrete wavelet transform of the bias-corrected log-smoothed periodogram of the batch means. The method proceeds by taking the inverse of the discrete wavelet transform of the thresholded wavelet coefficients to compute estimates of the batch means log-spectrum and  $\sigma^2$ . One of the real advantages of WASSP is that it incorporates a completely automated procedure for determining the data-truncation point that is highly effective in eliminating initialization bias; see Lada et al. (2007). Lada et al. (2006, 2007) present experimental comparisons between classical NBM methods, the ASAP3 method in Section 3.4 and WASSP.

### 3.7 Quantile Estimation

A variety of methods have been proposed for estimating quantiles of steady-state data (see Iglehart 1976; Seila 1982ab; Heidelberger and Lewis 1984; Chen and Kelton 2006). The methods differ in the way the variance of the sample quantile is estimated. It should be mentioned that quantile estimation is typically a harder problem than the estimation of steady-state means.

### 3.8 Density Estimation

In addition to point and interval estimates, users are often interested in estimating the density functions (or CDFs) of random variables generated by a computer simulation. Several simulation packages can generate histograms, but such plots are often “poor” estimates of the unknown density function because their shape depends heavily on the chosen origin and the bin width. Although the statistical literature contains many state-of-the-art density estimation techniques, such as those based on kernel functions, the simulation literature (in particular texts) barely mentions such techniques, and only within the context of independent input data. The book chapter by Alexopoulos (2006) attempts to close the gap between the statistical and simulation literatures by reviewing univariate kernel density estimators based on independent samples and sample paths of stationary dependent processes.

### 3.9 Multivariate Estimation

Frequently, the output from a single simulation run is used for estimating several system parameters. The estimators of these parameters are typically correlated. As an example, consider the average customer delays at two stations on a path of a queueing network. In general, Bonferroni’s

inequality can be used for computing a conservative confidence coefficient for a set of CIs. Indeed, suppose that  $D_i$  is a  $1 - \alpha$  CI for the parameter  $\mu_i$ ,  $i = 1, \dots, m$ . Then  $\Pr[\cap_{i=1}^m \{\mu_i \in D_i\}] \geq 1 - \sum_{i=1}^m \alpha_i$ .

This result can have serious implications as for  $m = 10$  and  $\alpha_i = 0.10$  the r.h.s. of the above inequality is equal to 0. If the overall confidence level must be at least  $1 - \alpha$ , then the  $\alpha_i$  can be chosen so that  $\sum_{i=1}^m \alpha_i = \alpha$ . Multivariate estimation methods are described in Charnes (1990, 1991) and Chen and Seila (1987).

## REFERENCES

- Aktaran-Kalaycı T., C. Alexopoulos, N. T. Argon, D. Goldsman, and J. R. Wilson. 2006. Exact expected values of variance estimators in steady-state simulation. *Naval Research Logistics Quarterly* 54(4):397-410.
- Alexopoulos, C. 2006. Statistical estimation in computer simulation. In *Handbooks in Operations Research and Management Science, Vol. 13*, eds. S. C. Henderson, and B. L. Nelson, 193–223. Amsterdam, Netherlands: Elsevier B. V.
- Alexopoulos, C., S. Andradóttir, N. T. Argon, and D. Goldsman. 2006a. Replicated batch means variance estimators in the presence of an initial transient. *ACM Transactions on Modeling and Computer Simulation* 16(4):317–328.
- 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* to appear. Available online via <ftp://ftp.ncsu.edu/pub/eos/pub/jwilson/ovestv72.pdf> [accessed June 17, 2007].
- Alexopoulos, C., and D. Goldsman. 2004. To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation* 14:76–114.
- Alexopoulos, C., D. Goldsman, and R. F. Serfozo. 2006b. Stationary processes: Statistical estimation. In *Encyclopedia of Statistical Sciences*, 2d Ed., eds. N. Balakrishnan, C. Read, and B. Vidakovic 7991–8006. Hoboken, New Jersey: John Wiley & Sons.
- Alexopoulos, C., and A. F. Seila. 1998. Output data analysis. In *Handbook of Simulation*, ed. J. Banks, Chapter 7. New York: John Wiley & Sons.
- Anderson, T. W. 1984. *An introduction to multivariate statistical analysis*. New York: John Wiley & Sons.
- Antonini, C., C. Alexopoulos, D. Goldsman, and J. R. Wilson. 2007. Area variance estimators for simulation using folded standardized time series. Technical report, H. Milton Stewart School of In-

- dustrial and systems Engineering, Georgia Institute of Technology, Atlanta, GA. Available online via <http://www.ise.ncsu.edu/jwilson/fold07iie.pdf> [accessed July 9, 2007].
- Argon, N. T., and S. Andradóttir. 2006. Replicated batch means for steady-state simulations. *Naval Research Logistics* 53(6):508–524.
- Billingsley, P. 1968. *Convergence of probability measures*. New York: John Wiley & Sons.
- Calvin, J. M. 2007. Simulation output analysis using integrated paths. *ACM Transactions on Modeling and Computer Simulation* 17(3) to appear.
- Calvin, J. M., and M. K. Nakayama. 2006. Permuted standardized time series estimators for steady-state simulation. *Mathematics of Operations Research* 31(2):351–368.
- Chang, B.-Y. 2004. Estimation techniques for nonlinear functions of the steady-state mean in computer simulation. Ph.D. Thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.
- Charnes, J. M. 1990. Power comparisons for the multivariate batch-means method. In *Proceedings of the 1990 Winter Simulation Conference*, 281–287. Piscataway, New Jersey: IEEE.
- Charnes, J. M. 1991. Multivariate simulation output analysis. In *Proceedings of the 1991 Winter Simulation Conference*, 187–193. Piscataway, New Jersey: IEEE.
- Chen, E. J., and W. D. Kelton. 2006. Quantile and tolerance-interval estimation in simulation. *European Journal of Operational Research* 168(2):520–540.
- Chen, R. D., and A. F. Seila. 1987. Multivariate inference in stationary simulation using batch means. In *Proceedings of the 1987 Winter Simulation Conference*, 302–304. Piscataway, New Jersey: IEEE.
- 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.-H., D. Goldsman, and B. Melamed. 1997. Large-sample results for batch means. *Management Science* 43:1288–1295.
- Chow, Y. S., and H. Robbins. 1965. On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Annals of Mathematical Statistics* 36:457–462.
- Conway, R. W. 1963. Some tactical problems in digital simulation. *Management Science* 10:47–61.
- Crane, M. A., and D. L. Iglehart. 1975. Simulating stable stochastic systems III: Regenerative processes and discrete-event simulations. *Operations Research* 23:33–45.
- Damerdj, H. 1994a. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* 19:494–512.
- Damerdj, H. 1994b. On the batch means and area variance estimators. In *Proceedings of the 1994 Winter Simulation Conference*, 340–344. Piscataway, New Jersey: IEEE.
- Damerdj, H. 1995. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research* 43:282–291.
- Durrett, R. 2005. *Probability: Theory and examples*, 3d Ed. Belmont, California: Brooks/Cole–Thomson Learning.
- Fishman, G. S. 1973. Statistical analysis for queueing simulations. *Management Science* 20:363–369.
- Fishman, G. S. 1974. Estimation of multiserver queueing simulations. *Operations Research* 22:72–78.
- Fishman, G. S. 1978. *Principles of discrete event simulation*. New York: John Wiley & Sons.
- Fishman, G. S. 2001. *Discrete-event simulation*. New York: Springer-Verlag.
- Fishman, G. S., and L. S. Yarberry. 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.
- Glynn, P. W., and D. L. Iglehart. 1985. Large-sample theory for standardized time series: An overview. In *Proceedings of the 1985 Winter Simulation Conference*, 129–134. Piscataway, New Jersey: IEEE.
- Glynn, P. W., and D. L. Iglehart. 1990. Simulation analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.
- Glynn, P. W., and W. Whitt. 1991. Estimating the asymptotic variance with batch means. *Operations Research Letters* 10:431–435.
- Goldsman, D., K. Kang, and A. F. Seila. 1999. Cramér-von Mises variance estimators for simulations. *Operations Research* 47, 299–309.
- Goldsman, D., M. Meketon, and L. W. Schruben. 1990. Properties of standardized time series weighted area variance estimators. *Management Science* 36:602–612.
- Goldsman, D., and L. W. Schruben. 1990. New confidence interval estimators using standardized time series. *Management Science* 36:393–397.
- Hald, A. 1952. *Statistical theory with engineering applications*. New York: John Wiley & Sons.
- Heidelberger, P., and P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Operations Research* 32:185–209.
- Heidelberger, P., and P. D. Welch. 1981. A spectral method for confidence interval generation and run length control

- in simulations. *Communications of the ACM* 24:233–245.
- Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31(6):1109–1144.
- Henderson, S. C., and B. L. Nelson, eds. 2006. *Handbooks in operations research and management science, Vol. 13*. Amsterdam, Netherlands: Elsevier B. V.
- Iglehart, D. L. 1975. Simulating stable stochastic systems, V: Comparison of ratio estimators. *Naval Research Logistics Quarterly* 22:553–565.
- Iglehart, D. L. 1976. Simulating stable stochastic systems, VI: Quantile estimation. *Journal of the ACM* 23:347–360.
- Lada, E. K., N. M. Steiger, and J. R. Wilson. 2006. Performance evaluation of recent procedures for steady-state simulation analysis. *IIE Transactions* 38:711–727.
- Lada, E. K., and J. R. Wilson. 2006. A wavelet-based spectral procedure for steady-state simulation analysis. *European Journal of Operations Research* 174:1769–1801.
- Lada, E. K., J. R. Wilson, N. M. Steiger, and J. A. Joines. 2007. Performance of a wavelet-based spectral procedure for steady-state simulation analysis. *INFORMS Journal on Computing* 19(2):150–160.
- Law, A. M. 2006. *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*, 227–230. Piscataway, New Jersey: IEEE.
- Muñoz, D. F., and P. W. Glynn. 1997. A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Science* 43(8):1121–1135.
- Philipp, W., and W. Stout. 1975. Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoirs of the American Mathematical Society* 2.
- Sargent, R. G., K. Kang, and D. Goldsman. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Operations Research* 40:898–913.
- Satterthwaite, F. E. 1941. Synthesis of variance. *Psychometrika* 6:309–316.
- Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.
- Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.
- Seila, A. F. 1982a. A batching approach to quantile estimation in regenerative simulations. *Management Science* 28:573–581.
- Seila, A. F. 1982b. Percentile estimation in discrete event simulation. *Simulation* 39:193–200.
- Song, W.-M. T. 1996. On the estimation of optimal batch sizes in the analysis of simulation output analysis. *European Journal of Operations Research* 88:304–309.
- Song, W.-M. T., 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: An batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation* 15(1):39–73.
- von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference and the variance. *Annals of Mathematical Statistics* 12:367–395.
- Welch, P. D. 1983. The statistical analysis of simulation results. In *The Computer Performance Modeling Handbook*, ed. S. Lavenberg, 268–328. New York: Academic Press.
- Welch, P. D. 1987. On the relationship between batch means, overlapping batch means and spectral estimation, In *Proceedings of the 1987 Winter Simulation Conference*, 320–323. Piscataway, New Jersey: IEEE.
- Wilson, J. R., and A. A. B. Pritsker. 1978a. A survey of research on the simulation startup problem. *Simulation* 31:55–58.
- Wilson, J. R., and A. A. B. Pritsker. 1978b. Evaluation of startup policies in simulation experiments. *Simulation* 31:79–89.

## AUTHOR BIOGRAPHY

**CHRISTOS ALEXOPOULOS** is an Associate Professor in the 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 2007. He is also the Simulation Department Editor of *IIE Transactions*. 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).