ADVANCED METHODS FOR SIMULATION OUTPUT ANALYSIS

Christos Alexopoulos

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

ABSTRACT

This paper reviews statistical methods for analyzing output data from computer simulations of single systems. In particular, it focuses on the problems of choosing initial conditions and estimating steady-state system parameters. The estimation techniques include the replication/deletion approach, the regenerative method, the batch means method, the standardized time series method, the autoregressive method, and the spectral estimation method.

1 INTRODUCTION

The primary purpose of most simulation studies is the approximation of certain system parameters with the objective of identifying parameter values that optimize some system performance measures. If some of the input processes driving a simulation are random, then the output data are also random and runs of the simulation program only give *estimates* of system performance characteristics. Unfortunately, a simulation run does not usually produce independent, identically distributed (i.i.d.) observations; therefore "classical" statistical techniques are not directly applicable to the analysis of simulation output.

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 focuses on the use of output data for estimating system performance measures.

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

1. Terminating simulations. The termination of a transient simulation is caused by the occurrence of an event E. An example is the simulation of a computer network until n jobs are completed.

2. 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 (Law and Kelton 1991). 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 terminating simulations. Section 3 reviews approaches for removing bias due to initial conditions in steady-state simulations. Section 4 presents techniques for point and interval estimation of steady-state parameters. Finally, section 5 contains conclusions and recommendations for additional studies by the interested reader.

2 TERMINATING SIMULATIONS

We start with the output analysis methodology for terminating simulations. Suppose that we simulate a system until n output data $X_1, X_2, \ldots, X_n$ are collected with the objective of estimating $\mu = \mathsf{E}(\overline{X}_n)$, where $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ 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 ith hour. Clearly, $\overline{X}_n$ is an unbiased estimator for μ. Unfortunately, the X_i's are generally dependent random variables making the estimation of the variance $\mathsf{Var}(\overline{X}_n)$ a nontrivial problem. In many queueing systems the X_i's are positively correlated making the familiar estimator $S^2(n)/n = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2/[n(n-1)]$ a highly biased estimator of $\mathsf{Var}(\overline{X}_n)$.

To overcome this problem, one can run k independent replications of the system simulation. Assume that run i produces the output data

$X_{i1}, X_{i2}, \ldots, X_{in}$. Then the sample means

$$Y_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$$

are i.i.d. random variables,

$$\overline{Y}_k = \frac{1}{k} \sum_{i=1}^{k} Y_i$$

is also an unbiased estimator of μ, and

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

is an unbiased estimator of $\text{Var}(\overline{X}_n)$. If in addition n and k are sufficiently large, an approximate $100(1-\alpha)$ percent confidence interval for μ is

$$\mu \in \overline{Y}_k \pm t_{k-1,1-\alpha/2}\sqrt{\hat{V}_R/k}, \qquad (1)$$

where $t_{d,\gamma}$ represents the γ-quantile of the t distribution with d degrees of freedom.

Law and Kelton (1991) review sequential procedures for determining the number of replications required to estimate μ with a fixed error or precision. Their procedure for obtaining an estimate with a relative error $|\overline{X}_n - \mu|/|\mu| \leq \gamma$ and a $100(1-\alpha)$ percent confidence interval has performed well for initial sample size $n_0 \geq 10$ and $\gamma \leq 0.15$. A well-known sequential procedure for constructing a $100(1-\alpha)$ percent confidence interval for μ with a small absolute error $|\overline{X}_n - \mu| \leq \beta$ is due to Chow and Robbins (1965) (see also Nadas 1969). Law (1980) observed that the procedure is very sensitive to the value of β.

The method of replications can also be used for estimating performance measures other than means. For example, suppose that we want to estimate the p-quantile, say y_p, of the maximum queue size in a single-server queueing system during a fixed time window. We run k independent replications, denote by Y_i the maximum observed queue length during replication i, and let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(k)}$ be the order statistics corresponding to the Y_i's. Then a point estimate for y_p is

$$\hat{y}_p = \begin{cases} Y_{(kp)} & \text{if } kp \text{ is an integer} \\ Y_{(\lfloor kp+1 \rfloor)} & \text{otherwise} \end{cases}$$

and a confidence interval for y_p is described in Welch (1983, pp. 287–288).

3   INITIALIZATION PROBLEMS FOR STEADY-STATE SIMULATIONS

One of the hardest problems in steady-state simulations is the removal of the *initialization bias*. Suppose

that $\{X_i : i \geq 1\}$ is a discrete-time output stochastic process from a single run of a steady-state simulation with initial conditions I and assume that, as $n \to \infty$, $P(X_n \leq x|I) \to P(X \leq x)$, where X is the corresponding steady-state random variable. We consider the estimation of the steady-state mean $\mu = \lim_{n\to\infty} E(X_n|I)$. The problem with the use of the estimator $\overline{X}_n$ for a finite n is that $E(X_n|I) \neq \mu$.

The most commonly used method for eliminating the bias of $\overline{X}_n$ identifies a index $1 \leq l \leq n-1$ and *truncates* the observations $X_1, \ldots, X_l$. Then the estimator

$$\overline{X}_{n,l} = \frac{1}{n-l} \sum_{i=l+1}^{n} X_i$$

is generally less biased than $\overline{X}_n$ because the initial conditions primarily affect data at the beginning of a run. Several procedures have been proposed for the detection of a cutoff index l (see Fishman 1972; Gafarian, Ancker, and Morisaku 1978; Kelton and Law 1983; Schruben 1982; Schruben, Singh, and Tierney 1983; Wilson and Pritsker 1978a,b). The procedure of Kelton (1989) uses a *pilot* run to estimate the steady-state distribution and starts a production run by sampling from the estimated distribution. More sophisticated truncation rules and initialization bias tests have recently been proposed by Chance and Schruben (1992), Goldsman, Schruben, and Swain (1993), and Ockerman (1995).

We now briefly discuss the graphical procedure of Welch (1981, 1983) which is popular due to its simplicity and generality. Another graphical method has been proposed by Fishman (1978a,b) in conjunction with the batch means method (see section 4.3). Welch's method uses k independent replications with the ith replication producing observations $X_{i1}, X_{i2}, \ldots, X_{in}$ and computes the averages

$$\overline{X}_j = \frac{1}{k} \sum_{i=1}^{k} X_{ij}, \quad j = 1, \ldots, n.$$

Then for a given *time window* w, the procedure plots the moving averages

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

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 is similar to the choice of an interval width for a histogram.

4 STEADY-STATE ANALYSIS

Several methods have been developed for the estimation of steady-state system parameters. Below we briefly review these methods and provide the interested reader with an extensive list of references. We focus on the estimation of the steady-state mean μ of a discrete-time output process $\{X_i : i \geq 1\}$. Analogous methods for analyzing continuous-time output data are described in a variety of texts (Bratley, Fox, and Schrage 1987; Fishman 1978b; Law and Kelton 1991). The process $\{X_i\}$ is called *strictly stationary* if the joint distribution of $X_{i+j_1}, X_{i+j_2}, \ldots, X_{i+j_k}$ is independent of i for all indices $j_1, j_2, \ldots, j_k$. If $\mathsf{E}(X_i) = \mu$, $\mathsf{Var}(X_i) < \infty$ for all i, and the $\mathsf{Cov}(X_i, X_{i+j}) = C_j$ is independent of i, then $\{X_i\}$ is called *covariance-stationary*.

4.1 The Replication/Deletion Approach

This approach runs k independent replications, each of length n observations, and uses the method of Welch (1981, 1983) to discard the first l observations from each run. One then uses the i.i.d. sample means

$$Y_i = \frac{1}{n-l} \sum_{j=l+1}^{n} X_{ij}$$

to compute point and interval estimators for the steady-state mean μ (see section 2). The method is characterized by its simplicity and generality and works well if n and k are sufficiently large.

4.2 The Regenerative Method

This method assumes the identification of time indices at which the process $\{X_i\}$ probabilistically *starts over* and uses these regeneration epochs for obtaining i.i.d. random variables which can be used for computing point and interval estimates for the mean μ. The method was proposed by Crane and Iglehart (1974a, 1974b, 1975) and Fishman (1973, 1974). A complete treatment of the regenerative method is given in Crane and Lemoine (1977). More precisely, assume that there are (random) time indices $1 \leq T_1 < T_2 < \cdots$ 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 $Y_i = \sum_{j=T_i}^{T_i+1} X_j$ and $Z_i = T_{i+1} - T_i$ for $i = 1, 2, \ldots$ and assume that $\mathsf{E}(Z_i) < \infty$. Then the mean μ is given by

$$\mu = \frac{\mathsf{E}(Y_1)}{\mathsf{E}(Z_1)}.$$

Now suppose that one simulates the process $\{X_i\}$ over n cycles and collects the observations $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$. Then

$$\hat{\mu} = \frac{\overline{Y}_n}{\overline{Z}_n}$$

is a strongly consistent, although typically biased for finite n, estimator of μ. Furthermore, confidence intervals for μ can be constructed by using the random variables $Y_i - \mu Z_i, i = 1, \ldots, n$ and the central limit theorem (see Iglehart 1975). For small sample sizes and bounded Y_i and Z_i, one can compute the confidence interval in Alexopoulos (1993) which provides superior coverage over confidence intervals based on the central limit theorem at the expense of increased width.

The regenerative method is difficult to apply in practice because the majority of simulations have either no regenerative points or very long cycle lengths. A class of systems the regenerative method has successfully been applied to is inventory systems.

4.3 The Batch Means Method

The method of batch means is frequently used to estimate the steady-state mean μ or the $\mathsf{Var}(\overline{X}_n)$ and owns its popularity to its simplicity and effectiveness. Basic references on the method are Conway (1963), Fishman (1978a,b), Law and Carson (1979), and Schmeiser (1982).

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

To motivate the method, assume temporarily that the process $\{X_i\}$ is covariance-stationary and split the data into k batches, each consisting of b observations (assume $n = kb$). The ith batch consists of the observations

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

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

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

Since the process is covariance-stationary, it can be shown that

$$
\begin{aligned}
C_l(b) &\equiv \mathsf{Cov}[Y_i(b), Y_{i+l}(b)] \\
&= \frac{1}{b} \sum_{j=1-b}^{b-1} \left(1 - \frac{|j|}{b}\right) C_{lb+j}
\end{aligned}
$$

is independent of i. Therefore, if the autocovariance function $\{C_j\}$ is such that $C_l(b) \to 0$ as b increases, we can identify a batch size b for which the batch means $Y_i(b)$, $i = 1, \ldots, k$ are approximately i.i.d. normally distributed. Then we form the grand batch mean

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

estimate the $\text{Var}[Y_i(b)]$ by

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

and compute the approximate $100(1-\alpha)$ percent confidence interval for μ

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

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

We now focus on *consistent estimation* batch means methods. These methods assume the existence of a parameter σ^2, often referred to as the variance of the process $\{X_i\}$, such that a central limit theorem holds

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \sigma N(0,1) \quad \text{as } n \to \infty \qquad (3)$$

and aim at constructing an asymptotically consistent estimator for σ^2. In the following, $N(0,1)$ denotes the standard normal distribution, "$\xrightarrow{d}$" denotes convergence in distribution, and "$\xrightarrow{a.s.}$" denotes convergence with probability 1 (see Billingsley 1968). Also, $\{W(t) : t \geq 0\}$ is the standard Brownian motion process. The limiting result (3) is implied under the following assumption.

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

$$n(\overline{X}_n - \mu) = \sigma W(n) + O(n^{1/2-\lambda}) \quad \text{w.p. } 1.$$

This assumption states that for almost all sample paths, the process $\{n(\overline{X}_n - \mu)/\sigma\}$ is close to a standard Brownian motion.

The ASA is not restrictive as it holds under relatively weak assumptions for stationary ϕ-mixing processes, Markov chains and regenerative processes. See Damerdji (1994a) for details. Here we should point out that the constant λ is closer to $1/2$ for processes having little autocorrelation while it closer to zero for processes with high autocorrelation.

The ASA also implies $\overline{X}_n \xrightarrow{a.s.} \mu$ as $n \to \infty$. Now consider sequences of batch sizes $\{b_n : n \geq 1\}$ and numbers of batches $\{k_n : n \geq 1\}$ such that $k_n b_n \leq n$ and $k_n b_n/n \to 1$ as $n \to \infty$. Assume that the following assumptions are satisfied.

(A.1) $b_n \to \infty$ and $k_n \to \infty$ monotonically as $n \to \infty$

(A.2) $b_n^{-1} n^{1-2\lambda} \ln n \to 0$ as $n \to \infty$

(A.3) There exists a finite positive integer a such that

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

Damerdji (1994a) shows that, as $n \to \infty$,

$$\hat{V}_B(n) = \frac{1}{k_n - 1} \sum_{i=1}^{k_n} (Y_i(b_n) - \overline{Y}_{k_n})^2 \xrightarrow{a.s.} \sigma^2 \qquad (4)$$

and

$$Z_{k_n} = \frac{\overline{X}_n - \mu}{\sqrt{\hat{V}_B(n)/k_n}} \xrightarrow{d} N(0,1). \qquad (5)$$

The last display implies that

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

is an asymptotically valid $100(1 - \alpha)$ percent confidence interval for μ. One can easily verify that the batching sequence $b_n = \lfloor n^\theta \rfloor$ satisfies assumptions (A.1)-(A.3) if $\theta \in (1 - 2\lambda, 1)$. Damerdji (1994b) displays additional limiting properties of the variance estimator $\hat{V}_B(n)$.

We now elaborate on the following two batching rules.

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

The Square Root (SQRT) Rule. Choose batch sizes b_n and number of batches k_n so that $lim_{n\to\infty} b_n/\sqrt{n} = 1$ and $lim_{n\to\infty} k_n/\sqrt{n} = 1$.

The FNB rule produces asymptotically valid confidence intervals for μ but does not yield an asymptotically consistent estimator for σ^2. The confidence interval has the form (2) with b in the expression for

$\hat{V}_B$ replaced by b_n. Since consistent estimation methods produce confidence intervals whose half-widths have minimal expectation and variance (Glynn and Iglehart 1990), the FNB rule will induce confidence intervals that tend to be wider than confidence intervals resulting from consistent estimation methods.

The SQRT rule results from a theorem in Chien (1989). Under some additional stationarity assumptions on $\{X_i\}$, the theorem implies that $Z_{k_n} \xrightarrow{d} N(0,1)$ and the convergence will be fastest if b_n and k_n grow proportionally to $\sqrt{n}$. Unfortunately, in practice the SQRT rule tends to seriously underestimate the $\text{Var}(\overline{X}_n)$ for fixed n.

Fishman and Yarberry (1993) studied the estimation of the steady-state mean customer waiting time in an M/M/1 queue. The coverage rates of the confidence intervals produced by the FNB rule converge much faster to the nominal coverage than the coverage rates of the confidence intervals resulting from the SQRT rule. Furthermore, this disparity grows with increasing traffic intensity.

With the contrasts between the FNB and SQRT rules in mind, Fishman and Yarberry (1993) proposed two strategies that dynamically shift between the two rules. Both strategies compute confidence intervals at times $n_j \approx n_0 2^j, j = 0, 1, \ldots, J$.

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

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

Both strategies LBATCH and ABATCH yield random sequences of batch sizes. These sequences imply convergence results analogous to (4) and (5) (Fishman and Yarberry 1993). Preliminary experiments indicate that the more conservative ABATCH strategy comes closer to the FNB rule's superior coverage rates with shorter confidence intervals. These features make it an attractive compromise between the "extreme" FNB and SQRT rules.

We should point out that a plot of the batch means is a very useful tool for checking the effects of initial conditions, non-normality of batch means, and existence of correlation between batch means.

An interesting variation of the traditional batch means method is the method of *overlapping* batch means proposed by Meketon and Schmeiser (1984).

For given batch size b, this method uses all $n - b + 1$ overlapping batches to estimate μ and $\text{Var}(\overline{X}_n)$. The first batch consists of observations $X_1, \ldots, X_b$, the second batch consists of $X_2, \ldots, X_{b+1}$, etc. Welch (1987) noted that both traditional batch means and overlapping batch means are special cases of spectral estimation (see section 4.6) at frequency 0 and, more importantly, suggested that overlapping batch means yield optimal variance reduction when one forms sub-batches within each batch and applies the method to the sub-batches. For example, a batch of size 64 is split into 4 sub-batches and the first (overlapping) batch consists of observations $X_1, \ldots, X_{64}$, the second consists of observations $X_{17}, \ldots, X_{80}$, etc.

Another article in this volume (Sherman 1995) reviews a variety of batching rules and tests their efficiency.

4.4 The Standardized Time Series Method

This method was proposed by Schruben (1983). The standardized time series is defined by

$$T_n(t) = \frac{\lfloor nt \rfloor (\overline{X}_n - \overline{X}_{\lfloor nt \rfloor})}{\sigma \sqrt{n}}, \quad 0 \le t \le 1$$

and, under some mild assumptions (e.g., strict stationarity and ϕ-mixing),

$$(\sqrt{n}(\overline{X}_n - \mu), \sigma T_n) \xrightarrow{d} (\sigma W(1), \sigma B),$$

where $\{B(t) : t \ge 0\}$ is the standard Brownian bridge process (see Billingsley 1968). Informally, $\{X_i\}$ is ϕ-mixing if X_i and X_{i+j} are approximately independent for large j.

If $A = \int_0^1 \sigma B(t)\,dt$ is the area under B, then the identity

$$\mathsf{E}(A^2) = \sigma^2/12$$

implies that σ^2 can be estimated by multiplying an estimator of $\mathsf{E}(A^2)$ by 12. Suppose that the data $X_1, \ldots, X_n$ are divided into k (contiguous) batches, each of size b. Then for sufficiently large n the random variables

$$A_i = \sum_{j=1}^{b} [(n+1)/2 - j] X_{(i-1)b+j}, \quad i = 1, \ldots, k$$

become approximately i.i.d. normal and an estimator of $\mathsf{E}(A^2)$ is

$$\hat{\mathsf{E}}(A^2) = \frac{1}{(b^3 - b)k} \sum_{i=1}^{k} A_i^2.$$

Hence an (approximate) $100(1-\alpha)$ percent confidence interval for μ is

$$\mu \in \overline{Y}_k \pm t_{k,1-\alpha/2}\sqrt{\hat{V}_T/n},$$

where

$$\hat{V}_T = 12\mathsf{E}(A^2).$$

The standardized time series method is easy to implement and has asymptotic advantages over the batch means method (see Goldsman and Schruben 1984). However, in practice it can require prohibitively long runs as noted by Sargent, Kang, and Goldsman (1992). The theoretical foundations of the method are given in Glynn and Iglehart (1990). Additional developments on the method, as well as other standardized time series estimators, are contained in Goldsman, Meketon, and Schruben (1990) and Goldsman and Schruben (1984, 1990). Finally, Damerdji (1994a) shows that under the assumption of strong approximation in section 4.3, batching sequences satisfying assumptions (A.1)-(A.3) yield asymptotically consistent estimators for the process variance σ^2.

4.5   The Autoregressive Method

This method was developed by Fishman (1978b) and assumes that the output process $\{X_i\}$ is covariance-stationary with mean μ and $\sum_{j=-\infty}^{\infty}|C_j| < \infty$, and can be represented by the autoregressive model of order p

$$\sum_{j=0}^{p} b_j(X_{i-j} - \mu) = \epsilon_i,$$

where $b_0 = 1$ and $\{\epsilon_i\}$ is a sequence of uncorrelated random variables with mean 0 and variance σ_ϵ^2. The procedure in Fishman (1978b) determines an order p and computes estimates $\hat{b}_j$ and $\hat{\sigma}_\epsilon^2$ of b_j and σ_ϵ^2 respectively. Then for large n an approximate $100(1-\alpha)$ percent confidence interval for μ is

$$\mu \in \overline{X}_n \pm t_{d,1-\alpha/2}\sqrt{\hat{V}_A/n},$$

where

$$\hat{V}_A = \frac{\hat{\sigma}_\epsilon^2}{n\left(\sum_{j=0}^{p}\hat{b}_j\right)^2}$$

and the degrees of freedom are computed from

$$d = \frac{n\sum_{j=0}^{p}\hat{b}_j}{2\sum_{j=0}^{p}(p-2j)\hat{b}_j}.$$

The major difficulty with the use of the autoregressive method is the validity of the autoregressive model. A generalization of the method was proposed by Schriber and Andrews (1984).

4.6   The Spectral Estimation Method

This method also assumes that the process $\{X_i\}$ is covariance-stationary. Under this assumption, the variance of $\overline{X}_n$ is given by

$$\mathsf{Var}(\overline{X}_n) = \frac{1}{n}\left(C_0 + 2\sum_{j=1}^{n-1}(1 - j/n)C_j\right).$$

The name of the method is due to the fact that if $\sum_{j=-\infty}^{\infty}|C_j| < \infty$, then $n\mathsf{Var}(\overline{X}_n) \to 2\pi g(0)$ as $n \to \infty$, where $g(\lambda)$ is the *spectrum* of the process at frequency λ and is defined by

$$g(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}C_j e^{-i\lambda j}, \quad |\lambda| \le \pi,$$

where $i = \sqrt{-1}$. Therefore, for large n the estimation of $\mathsf{Var}(\overline{X}_n)$ can be viewed as that of estimating $g(0)$. Estimators of this variance have the form

$$\hat{V}_S = \frac{1}{n}\left(\hat{C}_0 + 2\sum_{j=1}^{p-1}w_j\hat{C}_j\right),$$

where p and the *weights* are chosen for improving the properties of the variance estimator $\hat{V}_S$. The selection of these parameters is discussed in Fishman (1978b) and Law and Kelton (1984). Further discussions of spectral methods are given in Heidelberger and Welch (1981a,b, 1983).

4.7   Quantile Estimation

A variety of methods have been proposed for estimating quantiles of steady-state data (see Iglehart 1976; Moore 1980; Seila 1982a,b; Heidelberger and Lewis 1984). The methods differ in the way the variance of the sample quantile is estimated. It should be mentioned that quantile estimation is a harder problem than the estimation of steady-state means.

4.8   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 confidence intervals. Indeed, suppose that D_i is a $100(1-\alpha)$ percent confidence interval for the parameter μ_i, $i = 1, \ldots, k$. Then

$$P\left(\cap_{i=1}^{k}\{\mu_i \in D_i\}\right) \ge 1 - \sum_{i=1}^{k}\alpha_i.$$

This result can have serious implications as for $k = 10$ 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 $100(1 - \alpha)$ percent, then the α_i's can be chosen so that $\sum_{i=1}^{k} \alpha_i = \alpha$. The existing multivariate estimation methods include Charnes (1989, 1990, 1991) and Chen and Seila (1987).

5 CONCLUSIONS

The purpose of this paper is to alert the user on a variety of issues and methodologies related to the analysis of output data from a simulation of a single system. Several aspects of output analysis were left out such as comparison of systems, design of simulation experiments, and variance reduction methods. These subjects are treated in a variety of articles in this volume and in several texts including Bratley, Fox, and Schrage (1987), Fishman (1978b), Kleijnen (1974, 1975), and Law and Kelton (1991).

ACKNOWLEDGMENTS

REFERENCES

Alexopoulos, C. 1993. Distribution-free confidence intervals for conditional probabilities and ratios of expectations. *Management Science* 40(12):1748-1763.

Billingsley, P. 1968. *Convergence of probability measures*, Wiley, New York.

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A guide to simulation*, 2d Ed. Springer-Verlag, New York, New York.

Chance, F., and L. W. Schruben. 1992. Establishing a truncation point in simulation output. Technical Report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Charnes, J. M. 1989. Statistical analysis of multivariate discrete-event simulation output. Ph.D. Thesis, Department of Operations and Management Science, University of Minnesota, Minneapolis, Minnesota.

Charnes, J. M. 1990. Power comparisons for the multivariate batch-means method. In *Proceedings of the 1990 Winter Simulation Conference*, 281-287. IEEE, Piscataway, New Jersey.

Charnes, J. M. 1991. Multivariate simulation output analysis. In *Proceedings of the 1991 Winter Simulation Conference*, 187-193. IEEE, Piscataway, New Jersey.

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. IEEE, Piscataway, New Jersey.

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.

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. 1974a. Simulating stable stochastic systems, I: General multiserver queues. *Journal of the ACM* 21:103-113.

Crane, M. A., and D. L. Iglehart. 1974b. Simulating stable stochastic systems, II: Markov chains. *Journal of the ACM* 21:114-123.

Crane, M. A., and D. L. Iglehart. 1975. Simulating stable stochastic systems III: Regenerative processes and discrete-event simulations. *Operations Research* 23:33-45.

Crane, M. A., and A. J. Lemoine. 1977. *An introduction to the regenerative method for simulation analysis*, Springer-Verlag, New York, New York.

Damerdji, H. 1991. Strong consistency and other properties of the spectral variance estimator. *Management Science* 37:1424-1440.

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

Damerdji, H. 1994b. On the batch means and area variance estimators. In *Proceedings of the 1994 Winter Simulation Conference*, 340-344. IEEE, Piscataway, New Jersey.

Fishman, G. S. 1972. Bias considerations in simulation experiments. *Operations Research* 20:785-790.

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

Fishman, G. S. 1978b. *Principles of discrete event simulation*, John Wiley, New York, New York.

Fishman, G. S., and L. S. Yarberry. 1993. An implementation of the batch means method. Technical Report UNC/OR/TR/93-1, Department of Operations Research, University of North Carolina, Chapel Hill, North Carolina.

Gafarian, A.V., C. J. Ancker, and F. Morisaku. 1978. Evaluation of commonly used rules for detecting steady-state in computer simulation. *Naval Research Logistics Quarterly* 25:511-529.

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

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. 1984. Asymptotic properties of some confidence interval estimators for simulation output. *Management Science* 30:1217-1225.

Goldsman, D., and L. W. Schruben. 1990. New confidence interval estimators using standardized time series. *Management Science* 36:393-397.

Goldsman, D., L. W. Schruben, and J. J. Swain. 1993. Tests for transient means in simulated time series. Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Heidelberger, P., and P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Operations Research* 32:185-209.

Heidelberger, P., and P. D. Welch. 1981a. 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. 1981b. Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development* 25:860-876.

Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31:1109-1144.

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.

Iglehart, D. L. 1978. The regenerative method for simulation analysis. In *Current Trends in Programming Methodology, Vol. III*, eds. K. M. Chandy, and K. M. Yeh, 52-71. Prentice-Hall, Englewood Cliffs, New Jersey.

Kelton, W. D. 1989. Random initialization methods in simulation. *IIE Transactions* 21:355-367.

Kelton, W. D., and A. M. Law. 1983. A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly* 30:641-658.

Kleijnen, J. P. C. 1974. *Statistical techniques in simulation, part I*, Marcel Dekker, New York, New York.

Kleijnen, J. P. C. 1975. *Statistical techniques in simulation, part II*, Marcel Dekker, New York, New York.

Law, A. M. 1980. Statistical analysis of the output data from terminating simulations. *Naval Research Logistics Quarterly* 27:131-143.

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.

Law, A. M., and W. D. Kelton. 1984. Confidence intervals for steady-state simulations, I: A survey of fixed sample size procedures. *Operations Research* 32:1221-1239.

Law, A. M., and W. D. Kelton. 1991. *Simulation modeling and analysis*, 2d Ed. McGraw-Hill, New York.

Meketon, M. S., and B. W. Schmeiser. 1984. Overlapping batch means: Something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, 227-230. IEEE, Piscataway, New Jersey.

Moore, L. W. 1980. Quantile estimation in regenerative processes. Ph.D. Thesis, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, North Carolina.

Nadas, A. 1969. An extension of the theorem of Chow and Robbins on sequential confidence intervals for the mean. *Annals of Mathematical Statistics* 40:667-671.

Ockerman, D. H. 1995. Initialization bias tests for stationary stochastic processes based upon standardized time series techniques. Ph.D. Thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Sargent, R. G., K. Kang, and D. Goldsman. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Operations Research* 40:898-913.

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

Schmeiser, B. W., and W.-M. T. Song. 1993. Optimal mean-squared-error batch sizes. To appear in *Management Science*.

Schriber, T. J., and R. W. Andrews. 1984. An ARMA-based confidence interval for the analysis of simulation output. *American J. Math. Manage-*

ment Sci. 4:345–373.

Schruben, L. W. 1981. Control of initialization bias in multivariate simulation response. *Communications of the ACM* 24:246–252.

Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30:569–590.

Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.

Schruben, L. W., H. Singh, and L. Tierney. 1983. Optimal tests for initialization bias in simulation output. *Operations Research* 31:1167–1178.

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.

Sherman, M. 1995. On batch means in the simulation and statistical communities. In *Proceedings of the 1995 Winter Simulation Conference*, this volume. IEEE, Piscataway, New Jersey.

Welch, P. D. 1981. On the problem of the initial transient in steady state simulations. Technical Report, IBM Watson Research Center, Yorktown Heights, New York.

Welch, P. D. 1983. The statistical analysis of simulation results. In *The computer performance modeling handbook*, ed. S. Lavenberg, 268–328, Academic Press, New York.

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. IEEE, Piscataway, New Jersey.

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. He received his Ph.D. in Operations Research from the University of North Carolina at Chapel Hill. His research interests are in the areas of applied probability, statistics, and optimization of stochastic systems. His recent work involves problems related to the optimal design of telecommunications and transportation systems. Dr. Alexopoulos is a member of ORSA and the TIMS College of Simulation. He is an Associate Editor for *IIE Transactions* and Co-Editor for this volume.