

A CONCISE HISTORY OF SIMULATION OUTPUT ANALYSIS

Christos Alexopoulos

W. David Kelton

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

Department of Operations, Business Analytics,
and Information Systems
University of Cincinnati
Cincinnati, OH 45221-0130, USA

ABSTRACT

This paper offers a concise history of simulation output statistical analysis during the last six decades. Given the space limitations and historical perspective, we focus on the creation of the main concepts and methodologies that shaped the area, and proceed with a brief description of their developmental stages. We direct most of our attention to mean and quantile estimation, especially for steady-state simulations, since the bulk of the literature has been in this area, but briefly mention other topics like density estimation.

1 INTRODUCTION

Since the dawn of computer simulation in the late 1940s with the infancy of the Monte Carlo method (Metropolis and Ulam 1949), it has been recognized that appropriate statistical methods must be applied to the output of stochastic simulations in order to evaluate the performance of the system or process under study. The unfortunate lack of attention to this in too many applications may have been a main contributing factor to simulation's having acquired, early-on, the unwarranted but surprisingly sticky epithet "the method of last resort" (Harling 1958; Wagner 1969; Lucas et al. 2015).

Almost as early, some researchers (Conway et al. 1959; Conway 1963; Fishman 1967) started investigating more closely the statistical nature of simulation-generated output, and began offering constructive ideas on how to deal with the fact that these data tend to break the rules of classical statistics (such rules include being independent, identically distributed, normally distributed, and stochastically stationary). For example, successive times in queue (or in system) of parts exiting a manufacturing system are generally serially dependent (since a large, say, observation might tend to be followed next by another large observation), have distributions that cannot be negative (so cannot be normal), and tend to have initial trends, e.g., after a possibly-unrealistic starting condition like empty and idle. At the same time, simulation-generated output data offer advantages over traditional physically-collected data sets, especially in terms of sample size. Since then, this field has grown and matured, and has had increasing impact on the good practice of simulation via implementation of methods and recommendations in simulation software, promoting their everyday use in projects. In 1986 the Winter Simulation Conference program, under the leadership of Program Chair Stephen D. Roberts, split the "Methodology" track into two separate full tracks, Modeling Methodology and Analysis Methodology, the latter of which has since then focused on the statistical issues of interpreting simulation output data.

In this paper we attempt to recount some of the history of these achievements. This is a *history* paper, not a comprehensive survey of methods or an in-depth description of techniques, so we do not attempt to describe in detail all the advancements (or cite every paper) in the field, and we are aware that we have omitted various important developments (and papers). The eager reader may refer to a variety of book chapters and surveys (e.g., Pawlikowski 1990; Goldsman and Nelson 2006; Law 2015; Nelson 2013). We will focus on the statistical analysis of just a *single* system variant, or *scenario*, and will not discuss

related important areas like comparing/ranking alternative scenarios, optimum seeking, variance reduction, or design of simulation experiments; these topics are covered in other history papers in the *Proceedings* of this conference.

In Section 2 we provide some notation, and in Section 3 we briefly describe problems for finite-horizon simulations. Section 4, which contains several subsections, describes the more challenging situation for infinite-horizon, or *steady-state* simulations. Section 5 briefly concludes.

2 NOTATION

To conserve space, we limit the exposition to univariate discrete-time processes. Multivariate estimation methods were reviewed in Section 4 of Alexopoulos and Seila (1998). Assume that $\{X_1, X_2, \dots\}$ is a stationary output process from within a simulation run or replication, with marginal mean μ , variance σ_X^2 , and cumulative distribution function (CDF) $F(\cdot)$. For example, X_i could be the length of stay in a hospital emergency department of the i th exiting patient. Let $\{C_j : j = 0, \pm 1, \pm 2, \dots\}$ denote the autocovariance function of the process, with C_j being the covariance of observations at lag j . Also let \bar{X}_n be the sample mean of $\{X_1, \dots, X_n\}$ and let $S_n^2(X)$ be the respective sample variance. Under appropriate conditions (Alexopoulos et al. 2006), one can show that the sequence of functions $\xi_n(t) \equiv \lfloor nt \rfloor (\bar{X}_{\lfloor nt \rfloor} - \mu) / (\sigma \sqrt{n})$ for $t \in [0, 1]$ and $n = 1, 2, \dots$ satisfies the functional central limit theorem (FCLT)

$$\xi_n(\cdot) \xrightarrow[n \rightarrow \infty]{} \mathcal{W}(\cdot), \quad (1)$$

where $\lfloor \cdot \rfloor$ denotes the floor function, $\xrightarrow[n \rightarrow \infty]{} \mathcal{W}(\cdot)$ denotes weak convergence as $n \rightarrow \infty$, $\mathcal{W}(\cdot)$ is a standard Brownian motion on $[0, 1]$, and

$$\sigma^2 \equiv \lim_{n \rightarrow \infty} n \text{Var}(\bar{X}_n) = \sigma_X^2 + 2 \sum_{j=1}^{\infty} C_j < \infty$$

is the asymptotic variance of the process. Taking $t = 1$ yields the well-known central limit theorem (CLT) for stationary sequences. If the data are independent, one has $\sigma^2 = \sigma_X^2$ and the classical CLT.

Another process that leads to estimators for σ^2 is the standardized time series (STS)

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

(Schruben 1983). Under the FCLT assumption (1), it can be shown that

$$\left[\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}, T_n(\cdot) \right] \xrightarrow[n \rightarrow \infty]{} (\mathcal{W}(1), \mathcal{B}(\cdot)),$$

where $\mathcal{B}(\cdot)$ is a standard Brownian bridge on $[0, 1]$ that is independent of $\mathcal{W}(1)$. We will revisit the STS in Section 4.

Typically, the form of $100(1 - \alpha)\%$ confidence intervals (CIs) for the mean will be

$$\bar{X}_n \pm t_{v, 1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}, \quad (3)$$

where $\hat{\sigma}^2$ is an estimator for the asymptotic variance σ^2 and $t_{v, \delta}$ is the δ -quantile of Student's t distribution with v degrees of freedom (the value of v depends on the underlying method).

In addition to the steady-state mean μ , which measures central tendency, users should seek the estimation of marginal quantiles, namely, $x_p \equiv \inf\{x : F(x) \geq p\}$ for various values of $p \in (0, 1)$. If the marginal CDF is also differentiable, a more-challenging problem is the estimation of the density function $f(x) = F'(x)$. We will revisit these problems in Section 4. We will also denote the δ -quantile of the standard normal distribution with z_δ and the ceiling function with $\lceil \cdot \rceil$.

3 FINITE-HORIZON SIMULATIONS

If the time frame of the system being simulated is finite, as defined by specific starting and stopping conditions, then each simulation run, or *replication*, obviously will start and stop accordingly. For instance, a retail store opens at a certain time each day, closes the entrance doors later that day to block further arrivals, then operates until all customers present have left. In a simulation of such a system, one would define output metrics of interest, such as the average time customers spend in the store, the proportion of customers who spend more than an hour in the store, the time-average number of customers present, or the utilization of the checkout staff.

If Y_i is such an output metric from the i th (entire) replication, i.e., is an *aggregate* measure across a replication, then the Y_i are independent and identically distributed (IID) across multiple replications, and methods of classical statistical analysis apply, such as CIs. The approach to statistical analysis is just to make IID replications, with the number of replications being determined by precision requirements. Law (1980, 1983) provides further detail.

Historically, the focus has been primarily on aggregate output metrics, and means of such are certainly of interest in many cases. But other performance metrics might be more important, like tail probabilities of output-metric distributions, or quantiles of those distributions. Nelson (2008) develops histogram-like graphical devices called MORE plots that give a better prediction of the future, especially uncertainty and risk, than do mean-based analyses. More-advanced methods for estimating quantiles have been proposed by Avramidis and Wilson (1998) and Dong and Nakayama (2017), whereas density-estimation techniques based on IID data are reviewed by Alexopoulos (2006).

4 STEADY-STATE SIMULATIONS

In this case, we are interested in long-run, infinite-horizon performance metrics. Section 4.1 provides a historical perspective of methods for estimating the mean, whereas Section 4.2 reviews methods for estimating the steady-state marginal distribution or specific quantiles. The nature of this document makes inclusion of some material from papers published by the two authors unavoidable.

4.1 Estimation of the Steady-State Mean

We begin with methods for estimating the steady-state mean μ . Naturally, we emphasize the most recent methods. For older methods, see Law and Kelton (1982, 1984), who surveyed and compared available methods up to that time. The first two methods for estimating the steady-state mean have their roots in econometrics and engineering.

The Spectral Method

The method of *spectrum analysis* was first proposed by Fishman and Kiviat (1967) and Fishman (1969). The name of the method is due to the fact that the asymptotic variance can be written as $\sigma^2 = 2\pi g(0)$, where $g(\omega) \equiv (2\pi)^{-1} \sum_{j=-\infty}^{\infty} C_j \exp(-i\omega j)$, $-\pi \leq \omega \leq \pi$, is the spectrum of the process, and $i \equiv \sqrt{-1}$. Thus, for large n , we have $V_n \equiv \text{Var}(\bar{X}_n) \approx 2\pi g(0)/n$ and the estimation of σ^2 can be viewed as estimating the spectrum at zero frequency. The variance of \bar{X}_n can be estimated using the well-known expression

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

and the point estimates $\hat{\sigma}_X^2 = S_n^2(X)$ and $\hat{C}_j = (n-j)^{-1} \sum_{i=1}^{n-j} (X_i - \bar{X}_n)(X_{i+j} - \bar{X}_n)$ (for $j = 1, \dots, n-1$), but the latter estimates are serially correlated and biased (nearly zero for large lags j). This has led to the development of estimators for V_n of the form

$$\tilde{V}_n = \frac{1}{n} \left[\hat{\sigma}_X^2 + 2 \sum_{j=1}^{q-1} w_q(j) \hat{C}_j \right],$$

where q is a user-specified integer and the weights $w_q(\cdot)$ are designed to improve the sampling properties of \tilde{V}_n . An approximate $100(1 - \alpha)\%$ CI for μ is given by $\bar{X}_n \pm t_{v, 1-\alpha/2} \tilde{V}_n^{1/2}$, where v depends on n , q , and the weight function (Fishman 1969; Fishman 1973; Heidelberger and Welch 1981, 1983).

This method has deep engineering roots, but requires extensive intervention with regard to the choice of q and the weight function. Asymptotic properties of spectral estimators were obtained by Damerdjji (1991). Lada and Wilson (2006) developed a wavelet-based spectral method and a sequential procedure (WASSP), but this procedure was outperformed by methods based on batch means (Tafazzoli et al. 2011b).

The Autoregressive Method

This method, developed by Fishman (1971, 1973, 1978b) assumes that the output process can be expressed as a q -order autoregressive model $\sum_{j=0}^q a_j (X_{i-j} - \mu) = \varepsilon_j$, where $a_0 = 1$ and ε_j are IID with mean zero and variance σ_ε^2 . Then one can show that $\sigma^2 = \sigma_\varepsilon^2 / [\sum_{j=0}^q a_j]^2$. Fishman (1973) presented a method that obtains an estimate \hat{q} of the order, estimates \hat{a}_j of the coefficients a_j , and an estimate $\hat{\sigma}_\varepsilon^2$ of the white-noise variance σ_ε^2 . If we set $\hat{a} = 1 + \sum_{j=1}^{\hat{q}} \hat{a}_j$, a natural estimator for $\text{Var}(\bar{X}_n)$ is $\hat{V}_n = \hat{\sigma}_\varepsilon^2 / [n\hat{a}^2]$, and an approximate $100(1 - \alpha)\%$ CI for μ is $\bar{X}_n \pm t_{\hat{v}, 1-\alpha/2} \hat{V}_n^{1/2}$, where the estimate for the degrees of freedom is given by $\hat{v} = n\hat{a} / [2\sum_{j=0}^{\hat{q}} (\hat{q} - 2j)\hat{a}_j]$.

A major issue is the suitability of the autoregressive model for complex output processes. Yuan and Nelson (1994) revisited the method with an alternative technique for estimating the autoregressive order q and the degrees of freedom v .

The Regenerative Method

This method was, apparently, proposed simultaneously by Crane and Iglehart (1974a, 1974b) and Fishman (1974) (incidentally, all three papers appeared concurrently in the respective first issue of 1974). For a complete treatment, see Crane and Lemoine (1977). This method differs from other competitors because it does not yield a CI of the form (3). The method assumes the identification of time indices at which the process $\{X_i\}$ probabilistically *starts over* (regenerates) and uses these regeneration epochs for obtaining IID random pairs that are used to estimate the mean μ . As a result, cycles serve as independent replications and the point estimates are free of initialization bias.

More precisely, assume that there are (random) time indices $T_1 < T_2 < \dots$ 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 $Z_i = T_{i+1} - T_i$ for $i = 1, 2, \dots$ and assume that $E(Z_i) < \infty$. Under mild conditions, $\mu = E(Y_1) / E(Z_1)$.

Suppose that one simulates the process over b cycles and collects the IID vectors $\{(Y_i, Z_i) : i = 1, \dots, b\}$. Then $\hat{\mu}_b = \bar{Y}_b / \bar{Z}_b$ is a strongly consistent, although typically biased for finite b , estimator of μ . CIs for μ can be constructed by using the random variables $D_i = Y_i - \mu Z_i$, and the central limit theorem. Since $E(D_i) = 0$ and $\delta^2 \equiv \text{Var}(D_i) = \text{Var}(Y_i) - 2\mu \text{Cov}(Y_i, Z_i) + \mu^2 \text{Var}(Z_i)$, we have $\sqrt{b} \bar{D}_b / \delta \xrightarrow{b \rightarrow \infty} N(0, 1)$.

The classical, and most commonly used, approach estimates δ^2 by $S_b^2(D) = S_b^2(Y) - 2\hat{\mu}_b S_b(Y, Z) + \hat{\mu}_b^2 S_b^2(Z)$, where $S_b(Y, Z) = (b-1)^{-1} \sum_{i=1}^b (Y_i - \bar{Y}_b)(Z_i - \bar{Z}_b)$ is the sample covariance of Y_i and Z_i , to produce the approximate $100(1 - \alpha)\%$ CI $\hat{\mu}_b \pm z_{1-\alpha/2} S_b(D) / (\bar{Z}_b \sqrt{b})$.

Iglehart (1975) considered the situation where the cycle count b is small so that the asymptotic (in b) validity mentioned above may not hold. He defined the *jackknife* estimator of μ as $\tilde{\mu}_b = \sum_{i=1}^b \theta_i / b$, where

$$\theta_i = b \frac{\bar{Y}_b}{\bar{Z}_b} - (b-1) \frac{\sum_{j \neq i} Y_j}{\sum_{j \neq i} Z_j} \quad \text{for } i = 1, \dots, b,$$

and showed that it is first-order unbiased for μ . The corresponding CI $\tilde{\mu}_b \pm z_{1-\alpha/2} S_J / \sqrt{b}$, where $S_J^2 = (b-1)^{-1} \sum_{i=1}^b (\theta_i - \tilde{\mu}_b)^2$, typically has higher coverage probability than does its classical competitor. Alternative CIs based on *bootstrapping* were proposed by Choquet et al. (1999).

The regenerative method is difficult to apply in practice when one cannot identify regeneration states that are visited frequently. Two classes of systems to which the regenerative method has been successfully applied are inventory systems and highly reliable communications systems with repairs. An extension of this method is the semi-regenerative approach of Calvin et al. (2006).

4.1.1 Estimators Based on Nonoverlapping Batches

Nonoverlapping Batch Means (NBM)

The NBM method was mentioned in the seminal paper of Conway (1963) and has been the most robust method for computing CIs for the mean. The method divides a large sample $\{X_1, \dots, X_n\}$ into b nonoverlapping, contiguous batches, of size m each, and the batch means are the sample averages of the b batches, namely $Y_i = m^{-1} \sum_{k=1}^m X_{(i-1)m+k}$ for $i = 1, \dots, b$. Under the FCLT, one can show that for fixed b , the batch means become asymptotically independent as $m \rightarrow \infty$. This leads to the NBM estimator of the variance σ^2 ,

$$\hat{\sigma}_N^2(b, m) \equiv \frac{m}{b-1} \sum_{i=1}^b (Y_i - \bar{X}_n)^2, \quad (4)$$

is m times the sample variance of the batch means. Under mild assumptions, one has $\hat{\sigma}_N^2(b, m) \xrightarrow{m \rightarrow \infty} \sigma^2 \chi_{b-1}^2 / (b-1)$, a scaled chi-squared random variable with $b-1$ degrees of freedom (Glynn and Whitt 1991). Then a valid CI for μ is given in (3) with $v = b-1$. The method has also been extended to the estimation of a nonlinear function of the mean (Muñoz and Glynn 1997).

The fundamental problem is selection of an appropriate batch size m and batch count b . The primitive method of Mechanic and McKay (1966) searches within the set $\{16k, 64k, 256k, \dots, n/25\}$ (the default value is $k = 1$) and selects a value of m for which the batch means pass a test for independence. The method of Fishman (1978a) selects a batch size from the set $\{1, 2, 4, \dots, n/8\}$ such that the batch means pass von Neumann's randomness test (von Neumann 1941). Fishman (1978b) also suggested plotting the batch means in order to assess the potential effects of the initial transient, the independence of the batch means, and their departure from normality.

The frequently cited paper of Law and Carson (1979) showed that the batch means become uncorrelated as $m \rightarrow \infty$ under the milder assumption of weak stationarity for the output process and proposed a sequential procedure that yields a CI with a user-specified relative half-length. The procedure starts with 400 batches of size 2. Then it considers sample sizes that double every other iteration until an estimate of the lag-1 correlation among 400 batch means becomes smaller than 0.4 and larger than the estimated lag-1 correlation among 200 batch means. The procedure stops when the CI computed from 40 batch means satisfies the precision requirement. The major drawback of the methods between 1966 and 1979 is their ad hoc nature.

In a highly cited paper, Schmeiser (1982) analyzed the NBM method with regard to the mean squared error of the variance estimator and the coverage of the CI for the mean. The study concluded that selecting between 20 and 30 batches should suffice for the vast majority of experimental settings. Kelton and Law (1984) conducted a thorough evaluation of steady-state estimation procedures.

In a remarkable pair of papers, Damerджи (1994, 1995) established strong and mean-square consistency of the NBM variance estimator (4) as both $b, m \rightarrow \infty$ in an appropriate fashion and under a strong approximation assumption (ASA) that characterizes the rate of convergence in (1). At this juncture, we wish to point out that strong consistency leads to CIs whose half-length has smaller mean squared error than the half-length of CIs from the "cancellation" methods above. Damerджи's work motivated the iterative LABATCH.2 suite of algorithms of Fishman and Yarberrry (1997). Under the ASA, three of the four algorithms yield strongly consistent estimators for σ^2 as the sample size $n \rightarrow \infty$. In particular, each iteration of the more-robust ABATCH algorithm tests the batch means for independence using von Neumann's (1941) one-sided randomness test. If the batch means pass the test, the next iteration increases both the batch size m and batch count b by a factor of about $\sqrt{2}$; otherwise, the next iteration doubles the batch size and leaves the batch

count intact. All four algorithms have overall computational complexity $O(n)$ and storage requirements $O(\log_2 n)$, but are not truly sequential because they require extensive user intervention to identify an appropriate sample size for the CI that meets precision requirements.

To close the gap between the theoretical beauty of the LABATCH.2 methodology and minimal user intervention, James R. Wilson and various collaborators developed several NBM-based sequential procedures; the two procedures that merit special mention are ASAP3 by Steiger et al. (2005) and the Skart procedure of Tafazzoli and Wilson (2011). Both methods employ adjustments to account for residual autocorrelation or skewness in the sequence of batch means, limit the batch count, and employ conservative increases in the sample size during the latter stages; hence, they do not yield consistent variance estimators or provably asymptotically valid CIs as the CI precision requirement tightens. The trade-off between full theoretical rigor and overall performance in a plethora of challenging experimental settings has been the subject of extensive scientific debate. In our opinion, outstanding empirical performance is more important than theoretical “beauty,” in particular if sequential procedures are implemented in commercial software packages. Tafazzoli et al. (2011a) proposed a fixed-sample-size variant of N-Skart. The performance of a sequential procedure for finite samples can be quantified by means of the coverage function (Singham and Schruben 2012).

To Batch or Replicate?

The design of a steady-state simulation experiment involves the long-standing decision between batching observations in a single long run or replicating a number of shorter runs. The tradeoffs between the two approaches are well-known: batching ameliorates the effects of initialization bias, but yields statistics (e.g., batch means) that are typically correlated; replication yields independent statistics (e.g., sample means), but typically suffers from initialization bias at the onset of each replicate run; see Law (1977) for an early study.

Whitt (1991) studied this problem based on the efficiency of the estimator for μ (efficiency being defined as the product of the estimator’s MSE and the total simulation run length). He concluded that in the presence of initialization bias, one long run is typically more efficient when the autocorrelation function of the process dissipates faster compared to the rate at which the process approaches steady state.

The lengthy study of Alexopoulos and Goldsman (2004) compared two alternatives: one long run composed of b nonoverlapping batches of size m each versus b replications, each starting in the same state and having length m . The findings of this study, which includes substantial experimentation, are in line with those of Glynn and Heidelberger (1991), Whitt (1991), and Fishman (2001, Section 6.4), and are summarized as follows: If the simulation starts in steady state, batching and replication perform similarly with regard to the MSE of the point estimators of μ and σ^2 , but replication tends to perform better in terms of the coverage probability of the CI for μ . This victory for replication appears to be hollow, since in the presence of an initial transient, batching often outperforms replication, in particular with respect to the coverage probability of the CI for μ .

The frequently used replication/deletion method (Law 2015, Section 9.5.2) starts with the identification of a truncation index d such that the sample path $\{X_{d+1}, X_{d+2}, \dots\}$ is nearly stationary. This method performs b independent replications, with replication i generating the data set $\{X_{i,d+1}, \dots, X_{i,d+m}\}$, and uses the IID replicate averages $Y_i(d) \equiv m^{-1} \sum_{j=1}^m X_{i,d+j}$ to compute a CI for μ . The identification of an appropriate index d is a very challenging problem. Gafarian et al. (1978) investigated methods that existed at that time, but found that none of them worked well in practice. A different approach was undertaken by Wilson and Pritsker (1978a, 1978b). Kelton and Law (1983, 1984) assumed that the sequence of the means is monotone and proposed a method that worked well in a class of models. The graphical approach of Welch (1983) is frequently cited, but it can be data-wasteful when the process has an autocorrelation function with a pronounced tail (Alexopoulos and Seila 1998). An alternative approach based on a Kalman filter was the subject of Gallagher et al. (1996). The method of Kelton (1989) reduces truncation by randomizing the initial state. Statistical tests for transient detection were proposed by Schruben et al. (1983), Vassilacopoulos (1989), Yücesan (1993), and Goldsman et al. (1994).

Starting in 1997, K. Preston White and collaborators proposed several variants of the marginal standard error rule (MSER), cf. White (1997), White and Robinson (2010), and Hoad and Robinson (2011). The most cited variant (MSER-5) computes the batch means $Y_j = 5^{-1} \sum_{k=1}^5 X_{5(j-1)+k}$, for $j = 1, \dots, J \equiv \lfloor n/5 \rfloor$, and the truncated averages $\bar{Y}_{J,\ell} = (J - \ell)^{-1} \sum_{j=\ell+1}^J Y_j$ and selects the truncation index as

$$\ell^* \equiv \arg \min_{\ell \in \{0, 1, \dots, J-1\}} \frac{1}{(n - \ell)^2} \sum_{j=\ell+1}^J (Y_j - \bar{Y}_{J,\ell})^2, \quad \text{provided that } \ell^* \leq J/2.$$

In a recent paper, Wang and Glynn (2016) add a very interesting perspective into this problem by introducing ways for measuring the magnitude of the initial transient's effect in the context of a single, long replication. Specifically, they showed that, as the simulation run length tends to infinity, MSER's truncation point is determined by the minimizer of a certain random walk. This methodology was used to generate examples where at least one variant of MSER fails to predict accurately the duration of the initial transient. They concluded that the experimental testbed for transient truncation rules must be augmented. A couple of problematic cases are also presented in Law (2015, Examples 9.28 and 9.29).

Estimators Based on Standardized Time Series (STS)

Schruben (1983) devised the concept of estimating the variance σ^2 based on the STS (2). The formation of the weighted area estimator based on the sample $\{X_1, \dots, X_n\}$ starts with a function f that is continuous on $[0, 1]$ and normalized so that $\int_0^1 f(t) \mathcal{B}(t) dt \sim \sigma N(0, 1)$. Under some appropriate conditions, it can be shown that the square of the weighted area under the STS,

$$A(f; n) \equiv \left[\frac{1}{n} \sum_{i=1}^n f(i/n) \sigma T_n(i/n) \right]^2, \quad (5)$$

converges weakly to $\sigma^2 \chi_1^2$ (cf. Goldsman and Schruben 1990). The remarkable paper of Glynn and Iglehart (1990) proposed a theoretical framework for STS-based variance estimators, that included the NBM variance estimator as a special case, and the respective CIs. Alexopoulos et al. (2007b) review several variants of the weighted STS area estimator (5) corresponding to various weight functions f ; some are first-order unbiased (the multiplier of $1/n$ in the bias expression is zero), while the "orthonormal" estimators of Foley and Goldsman (1999) are also independent as $n \rightarrow \infty$.

An alternative class of estimators can be constructed by computing the weighted area under the square of the STS, $C(g; n) \equiv \sum_{i=1}^n g(i/n) [\sigma T_n(i/n)]^2 / n$, where the weight function g is bounded with a continuous second derivative on $[0, 1]$ and is normalized so that $E \left[\int_0^1 g(t) (\sigma \mathcal{B}(t))^2 dt \right] = \sigma^2$. Properties of the Cramér-von Mises (CvM) estimator $C(g; n)$ were studied by Goldsman et al. (1999).

Naturally, one can form b nonoverlapping batches, construct an STS-based estimator from each batch, and take the average of these estimators to obtain the respective batched variance estimator. For instance, if $A_1(f; m), A_2(f; m), \dots, A_b(f; m)$ are the STS area estimators from the b batches, then for fixed b , the random variables $A_i(f; m)$ become asymptotically independent as $m \rightarrow \infty$; hence the batched area estimator

$$\hat{\sigma}_A^2(f; b, m) \equiv \frac{1}{b} \sum_{i=1}^b A_i(f; m) \xrightarrow{m \rightarrow \infty} \sigma^2 \chi_b^2 / b$$

has the same mean but smaller variance than its constituents. A long-standing issue with batched and CvM variance estimators is the presence of small-sample bias. While some estimators are first-order unbiased, higher-order terms can be large when m is relatively small. This issue was explored in Aktaran-Kalaycı et al. (2007) as well as in the recent paper by Alexopoulos et al. (2016b); a potential resolution based on jackknifing was the subject of Dingerç et al. (2015).

Unfortunately, the development of multistage procedures based on STS estimators has not followed the frenetic pace of procedures based on the NBM estimator. The only known such procedures are the

two-stage procedure of Nakayama (1994) and the sequential procedure of Alexopoulos et al. (2016b). The latter method is conceptually simpler than its NBM-based competitors but uses a large initial batch size to overcome the aforementioned small-sample bias issue. Both procedures are based on area estimators only.

At this point, we wish to mention a set of papers that developed STS estimators based on transformations applied to the STS: Alexopoulos et al. (2010), Antonini et al. (2009), Calvin and Nakayama (2006), Calvin (2007, 2009), and Meterelliyo et al. (2015).

4.1.2 Estimators Based on Overlapping Batches

In a remarkable paper Meketon and Schmeiser (1984) proposed the method of batch means based on overlapping batches, with the full knowledge that the respective batch means are correlated. The seemingly bizarre method of overlapping batch means (OBM) uses results from spectral theory to obtain estimators for σ^2 that are superior to the NBM estimator, at least asymptotically.

Specifically, given the sample $\{X_1, \dots, X_n\}$ and a batch size of m observations, the first batch consists of the observations X_1, \dots, X_m , the second batch consists of X_2, \dots, X_{m+1} , etc. (overall, we form $n - m + 1$ overlapping batches). The respective batch means are $Y'_i \equiv m^{-1} \sum_{k=1}^{i+m-1} X_k$ for $i = 1, \dots, n - m + 1$. and the OBM estimator of σ^2 is

$$\hat{\sigma}_O^2(b, m) \equiv \frac{nm}{(n - m + 1)(n - m)} \sum_{i=1}^{n-m+1} (Y'_i - \bar{X}_n)^2, \quad (6)$$

where $b \equiv n/m$ is no longer the batch count.

Under mild moment and mixing conditions, Song and Schmeiser (1995) showed that the OBM estimator has a similar expected value to the NBM estimator, but $\text{Var}[\hat{\sigma}_O^2(b, m)]/\text{Var}[\hat{\sigma}_N^2(b, m)] \rightarrow 2/3$ as $b, m \rightarrow \infty$. Thus the OBM method yields better asymptotic performance than NBM with regard to mean squared error “for free.” Also, the behavior of $\text{Var}[\hat{\sigma}_O^2(b, m)]$ appears to be less sensitive to the choice of the batch size than does the behavior of $\text{Var}[\hat{\sigma}_N^2(b, m)]$ (see Table 1 in Song and Schmeiser 1995). An approximate $100(1 - \alpha)\%$ CI for the mean is given by (3) with degrees of freedom ν chosen so that $\text{Var}(\hat{\sigma}_O^2(b, m))$ is approximately $\sigma^2 \chi_\nu^2/\nu$, e.g., $\nu = 1.5(b - 1)[1 + (b - 1)^{-0.5-0.6b}]$ (Schmeiser 1986). For fixed b , the limiting distribution of the OBM estimator was obtained by Aktaran-Kalaycı et al. (2009), who also derived an approximation by a scaled χ^2 random variable.

Finally, Welch (1987) noted that both NBM and OBM estimators are special cases of the spectral estimator at frequency 0, and more importantly, suggested that overlapping batch means yield near-optimal variance reduction with partial (about 75%) overlapping. The sequential procedure of Alexopoulos et al. (2016b) used the maximum of two STS area estimators and the OBM estimator with 75% overlapping to obtain a more stable estimator for σ^2 .

Alexopoulos et al. (2007a, 2007b) published a pair of papers studying variance estimators based STS computed from overlapping batches. The derivation of these estimators is analogous to the OBM estimators. For example, one can compute area estimators from each of the overlapping batches, and then obtain the overlapping batched area estimator by averaging the $n - b + 1$ constituents to obtain the overlapping area estimator. This development addressed various challenges including the derivation of the limiting distributions as b remains constant and the batch size $m \rightarrow \infty$, the construction of computational procedures with $O(n)$ time complexity, and the derivation of practical approximations for the limiting distributions. A substantial experimental evaluation revealed that these overlapping STS estimators have substantially lower variance than their counterparts based on nonoverlapping batches (up to about 60%). The authors proceeded with development of variance estimators based on optimal linear combinations of overlapping variance estimators (Aktaran-Kalaycı et al. 2009).

4.2 Quantile and Density Estimation

Estimation of steady-state quantiles is substantially more challenging than is estimation of the mean. As with IID samples, the natural estimator of the p -quantile x_p based on the finite sample $\{X_1, \dots, X_n\}$ is $\hat{x}_p(n) = X_{(\lceil np \rceil)}$, where $X_{(i)}$ is the i th order statistic of the X_i values, i.e., the i th smallest of the X_i values. Define the indicator process $I_i(p) \equiv 1$ if $X_i \leq x_p$ or 0 otherwise, and let $\bar{I}_n(p) \equiv n^{-1} \sum_{i=1}^n I_i(p)$. For a sequence of random variables $\{\zeta_n : n \geq 1\}$ and a deterministic sequence $\{\vartheta_n : n \geq 1\}$, we say that $\zeta_n = O_{\text{a.s.}}(\vartheta_n)$ if there is a random variable B such that $|\zeta_n/\vartheta_n| \leq B < \infty$ with probability 1. Under a variety of conditions including mixing (Alexopoulos et al. 2016a; Deo 1973; Muñoz 2010; Sen 1972) one can obtain Bahadur's representation

$$\hat{x}_p(n) = x_p - \frac{\bar{I}_n(p) - p}{f(x_p)} + O_{\text{a.s.}} \left[\frac{(\log n)^{3/2}}{n^{3/4}} \right]$$

and the CLT

$$\frac{\sqrt{n}[\hat{x}_p(n) - x_p]}{\tau_p/f(x_p)} \xrightarrow[n \rightarrow \infty]{\text{d}} N(0, 1), \quad (7)$$

where $\tau_p^2 < \infty$ of the asymptotic variance of the process $\{I_i(p)\}$.

The regenerative setting offered the first natural window for estimating steady-state quantiles. The first fixed-sample-size methods were proposed by Iglehart (1976), Moore (1980), and Seila (1982a, 1982b). We give a high-level description of Seila's methodology because it uses batching. The method forms b batches of m cycles each. Then it sorts the data from the i th batch of cycles to obtain the quantile estimator $\hat{x}_p(i, m)$, and averages the batch quantile estimators to obtain the overall point estimator $\bar{x}_p(b, m) = b^{-1} \sum_{i=1}^b \hat{x}_p(i, m)$. The bias of this estimator can be reduced by means of jackknifing. Assume that m is even, and let $\hat{x}_p^{(1)}(i, m)$ and $\hat{x}_p^{(2)}(i, m)$ be point estimates computed from the first and second half, respectively, of the m cycles in the i th batch. Then the jackknife estimate from the i th batch is $\tilde{x}_{J,p}(i, m) = 2\hat{x}_p(i, m) - [\hat{x}_p^{(1)}(i, m) + \hat{x}_p^{(2)}(i, m)]/2$, the overall jackknife estimator for x_p is the average $\bar{x}_{J,p}(b, m)$ of the $\tilde{x}_{J,p}(i, m)$, and an approximate $100(1 - \alpha)\%$ CI for x_p is given by $\bar{x}_{J,p}(b, m) \pm t_{b-1, 1-\alpha/2} S_b(\tilde{x}_{J,p})/\sqrt{b}$, where $S_b^2(\tilde{x}_{J,p})$ is the sample variance of $\tilde{x}_{J,p}(i, m)$. Selection of b and m was left as an open problem. Seila recommended that m should be large enough so that the expected number of observations in each batch of m cycles is ≥ 100 and that $b \geq 10$.

The replication/deletion method is a natural candidate for estimating steady-state quantiles (Nelson 2008, Section 4). One simply uses the truncated data set $\{X_{i,d+1}, \dots, X_{i,d+m}\}$ from replication i to obtain a point estimate $\tilde{x}_p(i, d)$ and then uses the replicate quantile estimators to construct a CI for x_p . Unfortunately, this approach may suffer from the same symptoms as does its counterpart for estimating the mean.

The indirect method of Bekki et al. (2010) was designed to analyze cycle times in a production system that can be modeled as a queueing network. The method estimates a four-term Cornish-Fisher expansion (Cornish and Fisher 1937) for a selected quantile x_p based on the corresponding standard normal quantile z_p along with the first sample moments of the data $\{X_1, \dots, X_n\}$. As a result, the method can estimate multiple quantiles simultaneously without sorting data. Since sample moments computed from highly-correlated data can exhibit slow convergence to the true moments, the authors used sample sizes of 30 and 60 million to analyze systems with server utilization below 90% and above 90%, respectively.

The first sequential method for estimating steady-state quantiles was proposed by Raatikainen (1990). The method relies on nonoverlapping batches. Within each batch, quantile estimators are computed by the extended P^2 algorithm (Raatikainen 1987), which approximates the CDF $F(\cdot)$ by a piecewise-quadratic curve and then inverts it. The CI for x_p is constructed by means of the multivariate analogue of (7), spectral estimation of τ_p^2 using the method of Heidelberger and Welch (1981), and estimation of $f(x_p)$ based on the approximation of $F(x_p)$.

The sequential algorithms of Chen and Kelton (2006, 2008) are based on a small number of replications, typically set to 3. During the first run of the *zoom in* (ZI) algorithm, each iteration obtains bounds on the respective quantile estimate and discards data outside the bounds. The first run terminates based on several

rules, while the remaining runs use the bounds obtained from the first run; hence the outcomes of the replications are typically correlated. The *quasi-independent* (QI) algorithm attempts to create approximately independent observations within each run by taking progressively larger gaps between the data used to estimate the respective quantile. The two-phase QI algorithm of Chen and Kelton (2008) outperforms the original QI algorithm and also provides an estimate of the steady-state density $f(\cdot)$.

The Sequest method of Alexopoulos et al. (2016a) is an automated sequential procedure that delivers CIs for nonextreme quantiles ($0.05 \leq p \leq 0.95$) with user-specified relative precision. The algorithm takes advantage of ideas from recent batch-means-based methods (Tafazzoli and Wilson 2011) and sectioning (Asmussen and Glynn 2007, Section III.5a) and incorporates techniques to (a) reduce the bias in the point estimator due to the initial transient or inadequate run length; and (b) adjust the CI half-length to compensate for distorting effects due to autocorrelation or skewness in the quantile estimators computed from the nonoverlapping batches. For given n , the CI for x_p is centered at the point estimate $\hat{x}_p(n)$ computed from the entire sample and its half-length is computed from the quantile estimates $\hat{x}_p(1;m), \dots, \hat{x}_p(b;m)$ obtained from b batches of size m each.

The Sequem algorithm of Alexopoulos et al. (2017) addresses the estimation of extreme quantiles by grouping batches and extending the maximum transformation method of Heidelberger and Lewis (1984). An attractive byproduct of the grouping mechanism is the delivery of CIs for an arbitrary set of nonextreme quantiles at a marginal additional cost.

Wood and Schmeiser (1995) presented a method for estimating steady-state quantiles based on overlapping batches, without any follow-up archival publication.

In the realm of density estimation, Section 6 of Alexopoulos (2006) reviews kernel density estimators (KDEs) from stationary sequences. Two outstanding references with convergence properties of KDEs and their slow rates of convergence are Masry (1986) and Masry and Györfi (1987). For Markov chains, one can obtain the typical convergence rate of $n^{-1/2}$ using the look-ahead density-estimation method of Henderson and Glynn (2001).

5 CONCLUSIONS

We have presented a high-level history of methods for simulation output analysis. This area has been a cornerstone of the WSC Program since its inception (the first paper listed in the inaugural WSC *Proceedings* was by Fishman 1968), and the Conference has been the first outlet for many breakthroughs in analysis methodology. Given the plethora of over 1000 publications in this area that have appeared in archival journals and conference proceedings, we obviously can cite only a small subset. We chose to pass on the emerging thrust of quantifying the effects of simulation input uncertainty on estimates of output metrics (cf. Xie et al. 2014). This may cause some ill feelings, but the authors wish to assert that they tried their best under the circumstances and page limitations. Additional references are in several simulation texts and WSC tutorials on this subject spanning many years.

REFERENCES

- Aktaran-Kalaycı, T., C. Alexopoulos, N. T. Argon, D. Goldsman, and J. R. Wilson. 2007. "Exact expected values of variance estimators in steady-state simulation". *Naval Research Logistics* 54 (4): 397–410.
- Aktaran-Kalaycı, T., C. Alexopoulos, D. Goldsman, and J. R. Wilson. 2009. "Optimal linear combinations of overlapping variance estimators for steady-state simulation". In *Advancing the Frontiers of Simulation: A Festschrift in Honor of George Samuel Fishman*, edited by C. Alexopoulos, D. Goldsman, and J. R. Wilson, 291–328. New York: Springer Science+Business Media.
- Alexopoulos, C. 2006. "Statistical estimation in computer simulation". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, 193–223. Amsterdam: Elsevier B.V.

- Alexopoulos, C., C. Antonini, D. Goldsman, and M. Meterelloyoz. 2010. "Performance of folded variance estimators for simulation". *ACM Transactions on Modeling and Computer Simulation* 20 (3): Article 11.
- Alexopoulos, C., N. T. Argon, D. Goldsman, N. M. Steiger, G. Tokol, and J. R. Wilson. 2007a. "Efficient computation of overlapping variance estimators for simulation". *INFORMS Journal on Computing* 19 (3): 314–327.
- Alexopoulos, C., N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. 2007b. "Overlapping variance estimators for simulation". *Operations Research* 55 (6): 1090–1103.
- Alexopoulos, C., and D. Goldsman. 2004. "To batch or not to batch?". *ACM Transactions on Modeling and Computer Simulation* 14 (1): 76–114.
- Alexopoulos, C., D. Goldsman, A. C. Mokashi, K.-W. Tien, and J. R. Wilson. 2016a. "Sequest: A sequential procedure for estimating quantiles in steady-state simulations". *Operations Research*, in revision.
- Alexopoulos, C., D. Goldsman, A. C. Mokashi, and J. R. Wilson. 2017. "Automated simulation-based estimation of extreme steady-state quantiles via the maximum transformation". *ACM Transactions on Modeling and Computer Simulation* 27 (4): Article 22.
- Alexopoulos, C., D. Goldsman, and R. F. Serfozo. 2006. "Stationary processes: Statistical estimation". In *Encyclopedia of Statistical Sciences* (2nd ed.), edited by N. Balakrishnan, C. Read, and B. Vidakovic, Volume 12, 7991–8006. Hoboken, NJ: Wiley Interscience.
- Alexopoulos, C., D. Goldsman, P. Tang, and J. R. Wilson. 2016b. "SPSTS: A sequential procedure for estimating the steady-state mean using standardized time series". *IIE Transactions* 48:864–880.
- Alexopoulos, C., and A. F. Seila. 1998. "Output data analysis". In *Handbook of Simulation*, edited by J. Banks, 225–272. New York: John Wiley & Sons.
- Antonini, C., C. Alexopoulos, D. Goldsman, and J. R. Wilson. 2009. "Area variance estimators for simulation using folded standardized time series". *IIE Transactions* 41:134–144.
- Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer Science+Business Media.
- Avramidis, A. N., and J. R. Wilson. 1998. "Correlation-induction techniques for estimating quantiles in simulation experiments". *Operations Research* 46 (4): 574–591.
- Bekki, J. M., J. W. Fowler, G. T. Mackulak, and B. L. Nelson. 2010. "Indirect cycle time quantile estimation using the Cornish–Fisher expansion". *IIE Transactions* 42 (1): 31–44.
- Calvin, J. M. 2007. "Simulation output analysis using integrated paths". *ACM Transactions on Modeling and Computer Simulation* 17 (3): 13.
- Calvin, J. M. 2009. "Simulation output analysis using integrated paths II: Low bias estimators". *ACM Transactions on Modeling and Computer Simulation* 19 (3): Article 11.
- Calvin, J. M., P. W. Glynn, and M. K. Nakayama. 2006. "The semi-regenerative method of simulation output analysis". *ACM Transactions on Modeling and Computer Simulation* 16 (3): 280–315.
- 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.
- Chen, E. J., and W. D. Kelton. 2006. "Quantile and tolerance-interval estimation in simulation". *European Journal of Operational Research* 168:520–540.
- Chen, E. J., and W. D. Kelton. 2008. "Estimating steady-state distributions via simulation-generated histograms". *Computers and Operations Research* 35 (4): 1003–1016.
- Choquet, D., P. L'Ecuyer, and C. Léger. 1999. "Bootstrap confidence intervals for ratios of expectations". *ACM Transactions on Modeling and Computer Simulation* 9 (4): 326–348.
- Conway, R. W. 1963. "Some tactical problems in digital simulation". *Management Science* 10:47–61.
- Conway, R. W., B. M. Johnson, and W. L. Maxwell. 1959. "Some problems of digital simulation". *Management Science* 6:92–110.
- Cornish, E. A., and R. A. Fisher. 1937. "Moments and cumulants in the specification of distributions". *Revue de l'Institut International de Statistique* 5:307–320.

- 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 A. J. Lemoine. 1977. *An Introduction to the Regenerative Method for Simulation Analysis*. Berlin: Springer-Verlag.
- Damerdji, H. 1991. "Strong consistency and other properties of the spectral variance estimator". *Management Science* 37:1424–1440.
- Damerdji, H. 1994. "Strong consistency of the variance estimator in steady-state simulation output analysis". *Mathematics of Operations Research* 19:494–512.
- Damerdji, H. 1995. "Mean-square consistency of the variance estimator in steady-state simulation output analysis". *Operations Research* 43 (2): 282–291.
- Deo, C. M. 1973. "A note on empirical processes of strong mixing sequences". *Annals of Probability* 1 (5): 870–875.
- Dingeç, K. D., C. Alexopoulos, D. Goldsman, J. R. Wilson, W. Chiu, and T. Aktaran-Kalaycı. 2015. "Jackknifed variance estimators for simulation output analysis". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 459–471. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Dong, H., and M. K. Nakayama. 2017. "Quantile estimation with Latin hypercube sampling". *Operations Research* (65) 6: 1678–1695.
- Fishman, G. S. 1967. "Problems in the statistical analysis of simulation experiments: The comparison of means and the length of sample records". *Communications of the ACM* 10:94–99.
- Fishman, G. S. 1968. "Estimating reliability in simulation experiments". In *Proceedings of the 1968 Winter Simulation Conference*.
- Fishman, G. S. 1969. *Spectral Methods in Econometrics*. Cambridge, MA: Harvard University Press.
- Fishman, G. S. 1971. "Estimating sample size in computer simulation experiments". *Management Science* 18:21–38.
- Fishman, G. S. 1973. *Concepts and Methods in Discrete Event Digital Simulation*. New York: John Wiley & Sons.
- 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*. New York: John Wiley & Sons.
- Fishman, G. S. 2001. *Discrete-Event Simulation: Modeling, Programming, and Analysis*. New York: Springer-Verlag.
- Fishman, G. S., and P. J. Kiviat. 1967. "Spectral analysis of time series generated by simulation models". *Management Science* 13:525–557.
- Fishman, G. S., and L. S. Yarberrry. 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.
- Gafarian, A. V., C. J. Ancker, Jr., and F. Morisaku. 1978. "Evaluation of commonly used rules for detecting "steady-state" in computer simulation". *Naval Research Logistics Quarterly* 25:511–529.
- Gallagher, M. A., K. W. Bauer, Jr., and P. S. Maybeck. 1996. "Initial data truncation for univariate output of discrete-event simulations using the Kalman filter". *Management Science* 42:559–575.
- Glynn, P. W., and P. Heidelberger. 1991. "Analysis of initial transient deletion for replicated steady-state simulations". *Operations Research Letters* 10:437–443.
- Glynn, P. W., and D. L. Iglehart. 1990. "Simulation output 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., and B. L. Nelson. 2006. "Correlation-based methods for output analysis". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, Chapter 15. Oxford: Elsevier Science.
- 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. 1994. "Tests for transient means in simulated time series". *Naval Research Logistics* 41:171–187.
- Harling, J. 1958. "Simulation techniques in operations research—a review". *Operations Research* 6 (3): 307–319.
- 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* 24:233–245.
- Henderson, S. G., and P. W. Glynn. 2001. "Computing densities for Markov chains via simulation". *Mathematics of Operations Research* 26 (2): 375–400.
- Hoad, K., and S. Robinson. 2011. "Implementing MSER-5 in commercial simulation software and its wider implications". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasy, J. Himelspanch, K. P. White, and M. Fu, 495–503. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- 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 Association for Computing Machinery* 23:347–360.
- 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* 30:641–658.
- Kelton, W. D., and A. M. Law. 1984. "An analytical evaluation of alternative strategies in steady-state simulation". *Operations Research* 32:169–184.
- Lada, E. K., and J. R. Wilson. 2006. "A wavelet-based spectral procedure for steady-state simulation analysis". *European Journal of Operational Research* 174 (3): 1769–1801.
- Law, A. M. 1977. "Confidence intervals in discrete-event simulation: A comparison of replication and batch means". *Naval Research Logistics Quarterly* 24 (4): 667–678.
- Law, A. M. 1980. "Statistical analysis of the output data from terminating simulations". *Naval Research Logistics Quarterly* 27 (1): 131–143.
- Law, A. M. 1983. "Statistical analysis of simulation output data". *Operations Research* 31 (6): 983–1029.
- Law, A. M. 2015. *Simulation Modeling and Analysis*. 5th 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.
- Law, A. M., and W. D. Kelton. 1982. "Confidence intervals for steady-state simulations, II: A survey of sequential procedures". *Management Science* 28:550–562.
- 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.

- Lucas, T. W., W. D. Kelton, P. J. Sánchez, and S. M. Sanchez. 2015. "Changing the paradigm: Simulation, now a method of first resort". *Naval Research Logistics* 62 (4): 293–303.
- Masry, E. 1986. "Recursive probability density estimation for weakly dependent stationary processes". *IEEE Transactions on Information Theory* 32 (2): 254–267.
- Masry, E., and L. Györfi. 1987. "Strong consistency and rates for recursive probability density estimators of stationary processes". *Journal of Multivariate Analysis* 22:79–93.
- Mechanic, H., and W. McKay. 1966. "Confidence intervals for averages of dependent data in simulations II". Technical report ASDD, IBM Corporation, Yorktown Heights, NY.
- Meketon, M. S., and B. W. Schmeiser. 1984. "Overlapping batch means: Something for nothing?". In *Proceedings of the 1984 Winter Simulation Conference*, edited by S. Sheppard, U. W. Pooch, and C. D. Pegden, 227–230. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Meterelliyoz, M., C. Alexopoulos, D. Goldsman, and T. Aktaran-Kalaycı. 2015. "Reflected variance estimators for simulation". *IIE Transactions* 47 (11): 1185–1202.
- Metropolis, N., and S. Ulam. 1949. "The Monte Carlo method". *Journal of the American Statistical Association* 44 (247): 335–341.
- Moore, L. W. 1980. *Quantile Estimation Methods in Regenerative Processes*. Ph. D. thesis, Department of Statistics, University of North Carolina, Chapel Hill, NC.
- Muñoz, D. F. 2010. "On the validity of the batch quantile method for Markov chains". *Operations Research Letters* 38 (3): 223–226.
- 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* 43:1121–1135.
- Nakayama, M. K. 1994. "Two-stage stopping procedures based on standardized time series". *Management Science* 40 (9): 1189–1206.
- Nelson, B. L. 2008. "The MORE plot: Displaying measures of risk & error from simulation output". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 413–416. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation*. New York: Springer.
- Pawlikowski, K. 1990. "Steady-state simulation of queueing processes: A survey of problems and solutions". *ACM Computing Surveys* 22 (2): 123–170.
- Raatikainen, K. E. E. 1987. "Simultaneous estimation of several percentiles". *Simulation* 49:159–163.
- Raatikainen, K. E. E. 1990. "Sequential procedure for simultaneous estimation of several percentiles". *Transactions of the Society for Computer Simulation* 7 (1): 21–44.
- Schmeiser, B. W. 1982. "Batch size effects in the analysis of simulation output". *Operations Research* 30:556–568.
- Schmeiser, B. W. 1986. Personal communication.
- 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 (5): 573–581.
- Seila, A. F. 1982b. "Estimation of percentiles in discrete event simulation". *Simulation* 6:193–200.
- Sen, P. K. 1972. "On the Bahadur representation of sample quantiles for sequences of ϕ -mixing random variables". *Journal of Multivariate Analysis* 2 (1): 77–95.
- Singham, D. L., and L. W. Schruben. 2012. "Finite-sample performance of absolute precision stopping rules". *INFORMS Journal on Computing* 4:624–635.
- Song, W.-M., and B. W. Schmeiser. 1995. "Optimal mean-squared-error batch sizes". *Management Science* 41:110–123.

- Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2005. "ASAP3: A batch means procedure for steady-state simulation analysis". *ACM Transactions on Modeling and Computer Simulation* 15 (1): 39–73.
- Tafazzoli, A., N. M. Steiger, and J. R. Wilson. 2011a. "N-Skart: A nonsequential skewness- and autoregression-adjusted batch means procedure for simulation analysis". *IEEE Transactions on Automatic Control* 56 (2): 254–264.
- Tafazzoli, A., and J. R. Wilson. 2011. "Skart: A skewness- and autoregression-adjusted batch means procedure for simulation analysis". *IIE Transactions* 43 (2): 110–128.
- Tafazzoli, A., J. R. Wilson, E. K. Lada, and N. M. Steiger. 2011b. "Performance of Skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis". *INFORMS Journal on Computing* 23:297–314.
- Vassilacopoulos, C. 1989. "Testing for initialization bias in simulation output". *Simulation* 52:151–153.
- von Neumann, J. 1941. "Distribution of the ratio of the mean square successive difference to the variance". *Annals of Mathematical Statistics* 12 (4): 367–395.
- Wagner, H. M. 1969. *Principles of Operations Research With Applications to Managerial Decisions*. Englewood Cliffs, NJ: Prentice Hall.
- Wang, R. J., and P. W. Glynn. 2016. "On the marginal standard error rule and the testing of initial transient deletion methods". *ACM Transactions on Modeling and Computer Simulation* 27 (1).
- Welch, P. D. 1983. "The statistical analysis of simulation results". In *The Computer Performance Modeling Handbook*, edited by 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*, edited by A. Thesen, H. Grant, and W. D. Kelton, 320–323. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- White, K. P. 1997. "An effective truncation heuristic for bias reduction in simulation output". *Simulation* 69 (6): 323–334.
- White, K. P., and S. Robinson. 2010. "The problem of the initial transient (again), or why MSER works". *Journal of Simulation* 4:268–272.
- Whitt, W. 1991. "The efficiency of one long run versus independent replications in steady-state simulation". *Management Science* 37:645–666.
- Wilson, J. R., and A. A. B. Pritsker. 1978a. "Evaluation of startup policies in simulation experiments". *Simulation* 31 (3): 79–89.
- Wilson, J. R., and A. A. B. Pritsker. 1978b. "A survey of research on the simulation startup problem". *Simulation* 31 (2): 55–58.
- Wood, D. C., and B. W. Schmeiser. 1995. "Overlapping batch quantiles". In *Proceedings of the 1995 Winter Simulation Conference*, edited by C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 303–308. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian framework for quantifying uncertainty in stochastic simulation". *Operations Research* 62 (6): 1439–1452.
- Yuan, M., and B. L. Nelson. 1994. "Autoregressive-output-analysis methods revisited". *Annals of Operations Research* 53:391–418.
- Yücesan, E. 1993. "Randomization tests for initialization bias in simulation output". *Naval Research Logistics* 40 (5): 643–663.

AUTHOR BIOGRAPHIES

CHRISTOS ALEXOPOULOS is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests are in the areas of simulation, statistics, and optimization of stochastic systems. He is a member of INFORMS and an active participant in the Winter Simulation Conference, having been *Proceedings* Co-Editor in 1995, and Associate Program Chair in 2006. He served on the Board of Directors of WSC between

2008 and 2016. He is also an Associate Editor of the *ACM Transactions on Modeling and Computer Simulation*. His e-mail address is christos@gatech.edu, and his Web page is www.isye.gatech.edu/~christos.

W. DAVID KELTON is a Professor in the Department of Operations, Business Analytics and Information Systems at the University of Cincinnati, and a Visiting Professor of Operations Research at the Naval Postgraduate School. His research interests and publications are in the probabilistic and statistical aspects of simulation, applications of simulation, and stochastic models. He is co-author of the simulation textbooks *Simio and Simulation: Modeling, Analysis, and Applications* and *Simulation with Arena*, and was also co-author of the first three editions of *Simulation Modeling and Analysis*. He was Editor-in-Chief of the *INFORMS Journal on Computing* from 2000–2007, and was Simulation Area Editor for *Operations Research*. He served as WSC Program Chair in 1987, General Chair in 1991, was on the WSC Board of Directors from 1991–1999, and is a Founding Trustee of the WSC Foundation. He is a Fellow of both INFORMS and IIE. His e-mail address is david.kelton@uc.edu, and his Web page is www.cba.uc.edu/faculty/keltonwd/.