Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

IMPROVING STANDARDIZED TIME SERIES METHODS BY PERMUTING PATH SEGMENTS

James M. Calvin
Marvin K. Nakayama

Department of Computer Science
New Jersey Institute of Technology
Newark, New Jersey 07102 U.S.A.

ABSTRACT

We describe an extension procedure for constructing new standardized time series procedures from existing ones. The approach is based on averaging over sample paths obtained by permuting path segments. Analytical and empirical results indicate that permuting improves standardized time series methods. We also propose a new standardized time series method based on maximums.

1 INTRODUCTION

A basic problem in simulation is estimating the steady-state mean μ of a stochastic process. Under mild regularity conditions, it can be shown that for most processes having a steady state, time averages satisfy a strong law of large numbers and a central limit theorem. Thus, developing a point estimator for μ is straightforward: just run a long simulation and use the time average of the simulated process. However, constructing confidence intervals is more difficult in general because it is a non-trivial task to consistently estimate the variance constant σ in the relevant central limit theorem due to process autocorrelations. The regenerative method (Crane and Iglehart 1975) and spectral method (e.g., Anderson 1994) are two approaches to doing this, but each has certain drawbacks.

Schruben (1983) proposed the class of standardized time series (STS) methods as a way of constructing asymptotically valid confidence intervals for μ, with the advantage that these approaches do not require one to consistently estimate σ. Instead, σ is "cancelled out" in a manner reminiscent of the t-statistic. This approach is based on a function g, which, when applied to the entire sample path of the stochastic process, yields an estimate of the scale of the process (Glynn and Iglehart 1990). The validity of STS methods requires that the stochastic process satisfy a functional central limit theorem (e.g., Billingsley 1999), which states that when suitably scaled and centered, the stochastic process converges to a Brownian motion. STS methods can be extended by breaking up the sample path of the resulting Brownian motion into m batches, $m \geq 2$, applying the g function to each batch, and then combining the m g-values. We call this method *batching*. A number of different STS methods have been proposed in the literature, one example being the method of batch means.

In this paper, we introduce an approach to developing new STS methods from existing ones. Our idea is based on permuting segments of the sample path of the limiting Brownian motion. (This is similar to an approach we developed for regenerative processes in Calvin and Nakayama 1998, 2000.) Specifically, suppose that the sample path of the Brownian motion is divided into k equal-length segments, $k \geq 2$. Then permuting the segments and piecing them together yields another sample path. We apply an STS g function to each entire permuted sample path, and averaging over all permutations leads to our estimator. (Note that we differentiate between the terms *segments* and *batches*.) The half-width of the resulting confidence interval has, in the limit, less variability than that obtained by applying g to only the original sample path.

We demonstrate this approach by applying it to a specific example of a STS method, the (nonstandardized) maximum estimator, with $k = 2$ permuted segments. In this case, analytical calculations show that permuting 2 segments results in roughly a 25% reduction in the mean confidence-interval half width as compared to the nonpermuted maximum estimator. We also compare analytically the expected half widths when permuting 2 segments and batching with 2 batches, and show that permuting leads to a slight reduction in expected half width. Moreover, empirical results seem to show that permuting leads to confidence intervals with slightly better coverage than batching.

The (nonstandardized) maximum estimator we consider is apparently a new STS method. Schruben (1983) previously developed a similar approach, the standardized maximum, which is based on properties of the maximum of a Brownian bridge divided by a function of the location at which the maximum occurs. In contrast, the (nonstan-

dardized) maximum estimator only considers the value of the maximum.

The rest of the paper has the following organization. In Section 2 we review the general method of standardized time series. We introduce the (nonstandardized) maximum estimator in Section 3 and discuss in Section 4 the application of batching to the maximum estimator. In Section 5 we present the permuted STS approach and apply it our maximum estimator. Some empirical results are given in Section 6, and we conclude with some closing comments in Section 7.

2 BACKGROUND ON STANDARDIZED TIME SERIES

In this section we summarize some facts about standardized time series taken from Glynn and Iglehart (1990). Let $Y = \{Y(t) : t \geq 0\}$ be a real-valued stochastic process representing the output of a simulation experiment. For each $n \geq 1$, define the scaled process

$$Y_n(t) = \frac{1}{n}\int_{s=0}^{nt} Y(s)\,ds, \quad 0 \leq t \leq 1,$$

and note that $Y_n(1)$ is the sample mean of the process Y up to time n. Suppose that there is a real number μ and a positive number σ such that if we define processes $\{X_n(t) : 0 \leq t \leq 1\}$ by

$$X_n(t) = \sqrt{n}\,(Y_n(t) - \mu t)\,,$$

then

$$X_n \xrightarrow{\mathcal{D}} \sigma B \tag{1}$$

in $C([0,1])$ as $n \to \infty$, where $C([0,1])$ is the space of continuous real-valued functions on $[0,1]$, B is a standard Brownian motion, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Known as a functional central limit theorem (FCLT), (1) has been shown to hold under a variety of assumptions. Often the conditions require a type of asymptotic independence in the form of mixing conditions, which assert that two events far apart in time are almost independent (Billingsley 1999).

We are interested in constructing confidence intervals for the unknown parameter μ, which is the steady-state mean of Y. One way of accomplishing this is to apply a technique from the class of standardized time series methodologies. Each of these methods is based on a function $g : C([0,1]) \to \mathbb{R}$ from a class $\mathcal{M}$. The class $\mathcal{M}$ is defined as $\mathcal{M} = \{g = b \circ \Gamma : \ b \in \mathcal{N}\}$, where $\Gamma : C([0,1]) \to C([0,1])$ is given by

$$\Gamma(x)(t) = x(t) - tx(1)$$

and $\mathcal{N}$ is the class of functions $b : C([0,1]) \to \mathbb{R}$ satisfying the following three conditions:

(i) $b(\alpha x) = \alpha b(x)$ for $\alpha \in \mathbb{R}$ with $\alpha > 0$ and $x \in C([0,1])$,
(ii) $P\{(b \circ \Gamma)(B) > 0\} = 1$,
(iii) $P\{B \in D(b \circ \Gamma)\} = 0$, with $D(h)$ the set of discontinuities of a function h.

Each possible choice of $b \in \mathcal{N}$ and the resulting g function give rise to a standardized time series method. When $g \in \mathcal{M}$ and (1) holds,

$$\frac{Y_n(1) - \mu}{g(Y_n)} \xrightarrow{\mathcal{D}} \frac{B(1)}{g(B)}, \tag{2}$$

as $n \to \infty$, and $g(B)$ is independent of $B(1)$. Let $\Phi(x) = P(B(1) \leq x)$, $G(x) = P(g(B) \leq x)$, and $H(x) = P(B(1)/g(B) \leq x)$. Then

$$H(x) = \int_0^\infty \Phi(xy)G(dy).$$

By (2),

$$P\left(\frac{Y_n(1) - \mu}{g(Y_n)} \leq x\right) \to H(x)$$

as $n \to \infty$.

To construct confidence intervals, select $\gamma \equiv \gamma_g$ such that $H(\gamma) = 1 - \delta/2$. Then as the run length $n \to \infty$, the interval

$$\left[Y_n(1) - \gamma g(Y_n),\, Y_n(1) + \gamma g(Y_n)\right]$$

is an asymptotic $100(1-\delta)\%$ confidence interval for μ. The half-width of the above confidence interval is $L_n \equiv \gamma g(Y_n)$, and property (i) of class $\mathcal{N}$ implies that $L_n = \gamma g(X_n)/\sqrt{n}$. Thus, when (1) holds, properties (i) and (iii) of class $\mathcal{N}$ imply that $\sqrt{n}L_n \xrightarrow{\mathcal{D}} \gamma g(\sigma B) = \gamma\sigma g(B)$ as $n \to \infty$ by the continuous mapping theorem. Moreover, if $\{g(X_n) : n \geq 1\}$ is uniformly integrable, then $\sqrt{n}E[L_n] \to \gamma\sigma E[g(B)]$ as $n \to \infty$. Note that σ depends only on the original process Y and not on the choice of the function $g \in \mathcal{M}$, so we will consider $\psi(g) \equiv \gamma E[g(B)]$ as a measure of the limiting expected half-width of a confidence interval obtained using an STS method based on $g \in \mathcal{M}$.

3 STANDARDIZED TIME SERIES BASED ON MAXIMUM

We now introduce a new standardized time series methodology: the (nonstandardized) maximum estimator. Previously, Schruben (1983) developed the standardized maximum, which is based on properties of $\Gamma(B)(t^*)/(t^*(1-t^*))^{1/2}$,

where $t^* = \inf\{t \geq 0 : \Gamma(B)(t^*) = M\}$ and $M = \max\{\Gamma(B)(t) : 0 \leq t \leq 1\}$. (Observe that $\Gamma(B)$ is a Brownian bridge.) Our (nonstandardized) maximum estimator is based on properties of $\Gamma(B)(t^*)$.

Define the map $b : C([0,1]) \to \mathbb{R}$ by

$$b(x) = \max_{0 \leq t \leq 1} x(t). \tag{3}$$

It is straightforward to show that $b \in \mathcal{N}$, so if we let $g : C([0,1]) \to \mathbb{R}$ be defined by $g = b \circ \Gamma$, then $g \in \mathcal{M}$, which gives rise to an STS method. Calvin and Nakayama (2001) establish the following result.

Proposition 3.1. *When b is defined as in (3),*

$$H(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{4 + x^2}}\right).$$

The critical points of the distribution H in this case can be computed numerically. In particular, for a 90% confidence interval, $\gamma \approx 4.129$.

We now compare the nonstandardized and standardized maximum estimators in terms of the mean half-widths of the resulting confidence intervals. First since $\Gamma(B)$ is a Brownian bridge, when b is defined in (3), $P(b(\Gamma(B)) > y) = \exp\{-2y^2\}$ for $y \geq 0$; e.g., see p. 290 of Breiman (1992). Thus, for the nonstandardized maximum estimator, since $g(B) = b(\Gamma(B))$,

$$\begin{aligned} E[g(B)] &= \int_{y=0}^{\infty} P(g(B) > y)dy \\ &= \int_{y=0}^{\infty} e^{-2y^2}dy = \frac{\sqrt{2\pi}}{4}. \end{aligned} \tag{4}$$

Therefore, for a 90% confidence interval corresponding to the nonstandardized maximum estimator, our measure of the limiting expected half-width is $\psi(g) = 4.129 \cdot E[g(B)] \approx 2.588$.

For the standardized maximum method, the function $g = b \circ \Gamma$ is defined with $b(x) = x(t^*)/(t^*(1-t^*))^{1/2}$, where $t^* = \inf\{t \geq 0 : x(t) = M\}$ and $M = \max\{x(t) : 0 \leq t \leq 1\}$. In this case Schruben (1983) shows that $B(1)/g(B) \stackrel{\mathcal{D}}{=} t_3/\sqrt{3}$, where t_d is a Student-t random variable with d degrees of freedom and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Thus, it can be shown that $E[g(B)] = 2\sqrt{2/\pi}$ and the corresponding critical point for a 90% confidence interval is $2.353/\sqrt{3}$, so our measure of the limiting expected mean confidence interval half-width is $\psi(g) = (2.353/\sqrt{3})2\sqrt{2/\pi} \approx 2.168$. Consequently, the mean half-width corresponding to the nonstandardized maximum estimator is asymptotically about 19% larger than that for the standardized maximum. However, our empirical results in Section 6 seem to indicate that the coverage of confidence intervals constructed using the nonstandardized maximum method are closer to the nominal level than those for the standardized maximum.

4 BATCHING

We now describe a standard way to extend standardized time series methods to multiple "batches." The idea is rather than approximate the entire sample path by a single Brownian motion, we break up the sample path into m (non-overlapping) batches and approximate each batch by a Brownian motion. The function g is then applied to each batch, and we combine the m g-function values to come up with the overall estimator. We consider here the case of $m = 2$ batches, with the resulting g function denoted by g_2, but the idea holds for any arbitrary $m \geq 1$. This approach is discussed in Schruben (1983) and Glynn and Iglehart (1990).

Define maps Λ_i on $C([0,1])$ by

$$(\Lambda_i(x))(t) = x\left(\frac{i+t}{2}\right) - x\left(\frac{i}{2}\right)$$

for $i = 0, 1$, and set

$$g_2 = \left(b^2 \circ \Gamma \circ \Lambda_0 + b^2 \circ \Gamma \circ \Lambda_1\right)^{1/2} \tag{5}$$

for any $b \in \mathcal{N}$, where $h^2(x) = h(x)h(x)$ for a function h. It can be shown that $g_2 \in \mathcal{M}$ (Glynn and Iglehart 1990), so it corresponds to an STS method.

For the case of the nonstandardized maximum with 2 batches, Calvin and Nakayama (2001) show the following.

Proposition 4.1. *If b is defined as in (3) and batching is applied with $m = 2$ batches, then*

$$H(x) = \frac{1}{2} + \frac{x}{2\sqrt{x^2 + 8}}\left(1 + \frac{4}{x^2 + 8}\right).$$

For a 90% confidence interval, the corresponding critical point is $\gamma \approx 3.015$, which we computed numerically. Moreover,

$$E[g_2(B)] = \frac{3\sqrt{\pi}}{8}; \tag{6}$$

see Calvin and Nakayama (2001) for details.

5 PERMUTED STANDARDIZED TIME SERIES

We now present a new method for developing STS methods from existing ones. The basic idea entails dividing the sample path into k non-overlapping equal-length segments. Permute the k segments to generate another sample path, and

apply a g-function to the entire permuted path. Averaging over all permutations results in the permuted estimator.

We now give details for the special case of $k = 2$ segments. We permute the two halves of the path B, and call the permuted path $\widetilde{B}$. Specifically, let $B = \{B(t) : 0 \le t \le 1\}$, and define $\widetilde{B} = \{\widetilde{B}(t) : 0 \le t \le 1\}$ such that

$$\widetilde{B}(t) = \begin{cases} B(t+\frac{1}{2}) - B(\frac{1}{2}) & \text{if } 0 \le t \le \frac{1}{2} \\ B(t-\frac{1}{2}) + B(1) - B(\frac{1}{2}) & \text{if } \frac{1}{2} < t \le 1 \end{cases}.$$

Since increments of Brownian motion are stationary and independent, B and $\widetilde{B}$ have the same distribution. For any $g \in \mathcal{M}$, let

$$\widetilde{g}(B) = (g(B) + g(\widetilde{B}))/2. \tag{7}$$

Calvin and Nakayama (2001) establish the following theorem:

Theorem 5.1. *If $\widetilde{g}(B)$ is as defined in (7) with $g \in \mathcal{M}$, then*

(i) $E[\widetilde{g}(B)] = E[g(B)]$,
(ii) $Var[\widetilde{g}(B)] \le Var[g(B)]$.

Thus, asymptotically, permuting results in confidence intervals having less variability.

5.1 Permuted Maximum Estimator

We now apply permutations to the maximum estimator discussed in Section 3. Calvin and Nakayama (2001) prove the following:

Proposition 5.1. *Suppose b is defined as in (3). Also, suppose we apply permuting with $k = 2$ segments and no batching. Then*

$$H(x) = \frac{1}{2} + \frac{1}{\pi} x \sqrt{\frac{8}{8+x^2}} \tan^{-1}\left(\sqrt{\frac{8}{8+x^2}}\right) + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{4}\right) - \frac{1}{\pi}\frac{4x}{16+x^2}.$$

We now compare some properties of the non-batched (nonstandardized) maximum estimator with and without permutations. For 90% confidence intervals, the required critical point of the distribution H is $\gamma \approx 3.095$ for $\widetilde{g}$ (i.e., permuting), which we computed numerically. We previously saw that $\gamma \approx 4.129$ for g (i.e., no permuting). Since $E[g(B)] = E[\widetilde{g}(B)]$ as shown in Theorem 5.1(i), the mean confidence interval half-width is asymptotically about $3/4$ as large for the permuted estimator. The variability of the half-width is also reduced (Theorem 5.1(ii)).

We now summarize our results for the nonstandardized maximum estimator. Based on 90% confidence intervals, our measure of the limiting mean half-width for the non-permuted, non-batched estimator is

$$\psi(g) = 4.129 \cdot \frac{\sqrt{2\pi}}{4} \approx 2.588$$

by (4); for the permuted estimator ($k = 2$ segments) it is

$$\psi(\widetilde{g}) = 3.095 \cdot \frac{\sqrt{2\pi}}{4} \approx 1.940;$$

and for the batched estimator ($m = 2$ batches) it is

$$\psi(g_2) = 3.015 \cdot \frac{3\sqrt{\pi}}{8} \approx 2.004$$

by (6). Thus, in terms of limiting expected half-widths, permuting is better than not permuting, and permuting with 2 segments slightly outperforms batching with 2 batches. In the next section we empirically compare coverages of confidence intervals.

6 NUMERICAL EXPERIMENTS

We ran simulations on three different models to study the coverages of 90% confidence intervals based on the methods discussed in this paper. For each of the first two models, we considered two methods: the nonstandardized maximum using only permuting and only batching. The permuted estimators were constructed using $k = 2$ segments and the batched estimator with $m = 2$ batches. For the last model, we also examined the standardized maximum and the nonstandardized maximum, both of them with no permuting and no batching. For each model we estimated coverages for the constructed confidence intervals by running 100,000 independent replications.

In the first example we simulated the Ehrenfest urn model with 10 states. Figure 1 plots the observed coverage as a function of the simulation run length of each replication. Observe that, except for the shortest run lengths, the coverage for the permuted estimator is consistently higher (closer to the nominal level) than that for the batched estimator.

For the second numerical example we simulated a geometric jump model, which is a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ and transition probabilities

$$P_{0,i} = (1-p)p^i, \qquad P_{i+1,i} = 1.$$

In the example, $p = 1 - e^{-1/10}$. Figure 2 shows the coverage for the permuted estimator is consistently above that for the batched estimator.

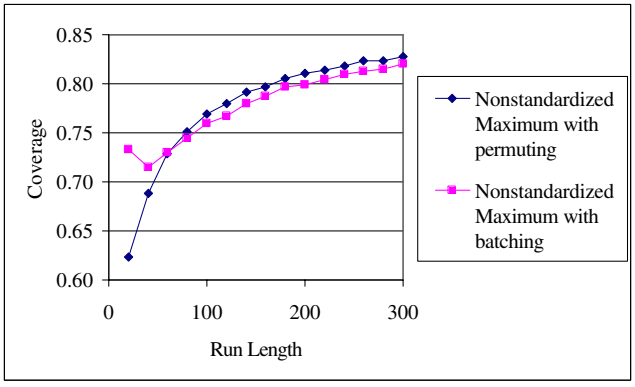

Figure 1: Observed Coverage Rates for Ehrenfest Urn Model

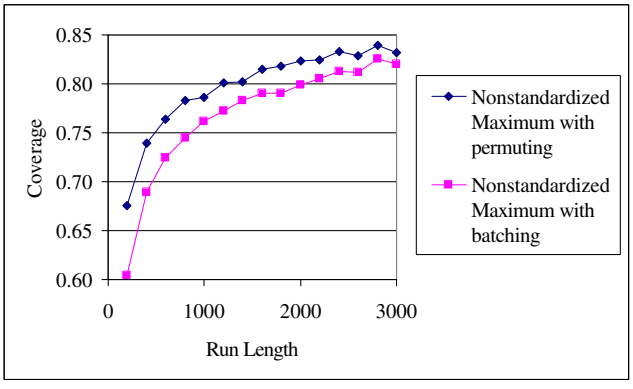

Figure 2: Observed Coverage Rates for Geometric Jump Model

The third model that we simulated is the embedded discrete-time chain of the number of customers in an M/M/1 queue on a truncated state space $S = \{0, 1, 2, \ldots, 100\}$. We set the arrival rate $\lambda = 0.8$ and service rate $\mu = 1.0$, so the traffic intensity is $\rho = 0.8$.

Figure 3 shows the average half-widths of the constructed 90% confidence intervals using the nonstandardized maximum method in three implementations: with no permuting and no batching, with only permuting, and with only batching. Also, the average half-widths are plotted for the standardized maximum estimator with no permuting and no batching. As previously shown in our theory for the nonstandardized maximum method in Section 5.1, permuting outperforms no permuting and no batching, and similarly, batching is better than no permuting and no batching. The difference between only permuting and only batching is small. Also, we see that when no permuting or batching is used, the standardized maximum method leads to smaller average half-widths than the nonstandardized method, which agrees with our theory in Section 3.

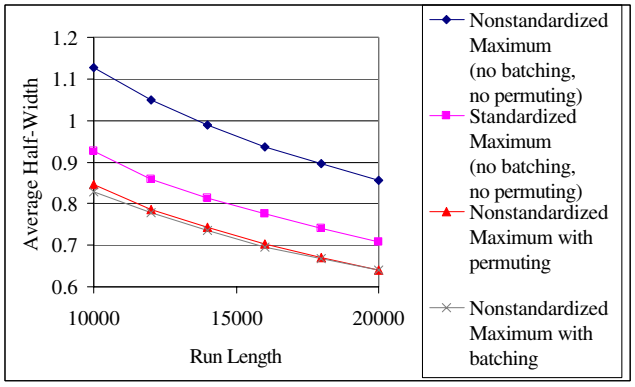

Figure 3: Average Confidence Interval Half-Widths for M/M/1 Queue

Figure 4 shows the observed coverages of confidence intervals for the M/M/1 model. Note that for the nonstandardized maximum method, permuting outperforms batching, which is consistent with what we saw for the other two models. Also, when no batching nor permuting is applied, coverages are higher for the nonstandardized maximum than for the standardized maximum. Thus, although the standardized maximum yields shorter confidence intervals on average, the coverage of these intervals is not as good.

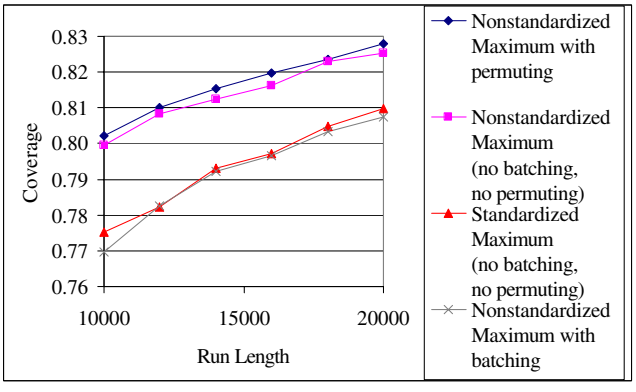

Figure 4: Observed Coverage Rates for M/M/1 Queue

In all of our experiments, permuting outperforms batching in terms of coverage. This may be due to the fact that when permuting, we apply the g-function to each entire permuted sample path, whereas batching applies the g-function to each of the smaller batches individually. Thus, under batching, the g-function is applied to shorter pieces of the sample path, which requires that the Brownian approximation from the FCLT holds for each of the smaller batches. But for permuting, since g is applied to each entire permuted path, the FCLT approximation is based on a larger sample.

7 CONCLUSIONS

In this paper we showed how to construct new STS methods from existing ones by permuting. Based on analytical and empirical studies, it appears that permuting can improve STS methods. In addition, permuting seems to have better small-sample properties than batching.

We also presented a new STS method, the nonstandardized maximum estimator. Compared to the standardized maximum, the nonstandardized maximum method appears to yield confidence intervals with better coverage but at the expense of larger mean half-widths.

ACKNOWLEDGMENTS

The work of the first author was supported in part by the National Science Foundation under grant DMI-9900117. The work of the second author was supported in part by the National Science Foundation under grants DMI-9624469 and DMI-9900117.

REFERENCES

Anderson, T. W. 1994. *The statistical analysis of time series*. New York : Wiley.

Billingsley, P. 1999. *Convergence of probability measures*, second edition. New York: John Wiley.

Breiman, L. 1992. *Probability*. Philadelpha: SIAM.

Calvin, J. M. and M. K. Nakayama. 1998. Using permutations in regenerative simulations to reduce variance. *ACM Transactions on Modeling and Computer Simulation* 8: 153–193.

Calvin, J. M. and M. K. Nakayama. 2000. Central limit theorems for permuted regenerative estimators. *Operations Research*, 48: 776–787.

Calvin, J. M. and M. K. Nakayama. 2001. Permuted standardized time series. Forthcoming.

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

Glynn, P. W. and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 14:1–16.

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

AUTHOR BIOGRAPHIES

JAMES M. CALVIN is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in Operations Research from Stanford University. Besides simulation output analysis, his research interests include global optimization and probabilistic analysis of algorithms.

MARVIN K. NAKAYAMA is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in Operations Research from Stanford University. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. He is an associate editor for *ACM Transactions on Modeling and Computer Simulation* and for *INFORMS Journal on Computing*. His research interests include applied probability, statistics, simulation-output analysis, variance-reduction techniques, gradient estimation, rare-event simulation, and random-variate generation.