Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

SEQUENTIAL ESTIMATION OF STEADY-STATE QUANTILES: LESSONS LEARNED AND FUTURE DIRECTIONS

Christos Alexopoulos
David Goldsman

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

Anup C. Mokashi

SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513-8617, USA

James R. Wilson

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

ABSTRACT

We survey recent developments concerning Sequest and Sequem, two simulation-based sequential procedures for estimating steady-state quantiles. These procedures deliver improved point and confidence-interval (CI) estimators of a selected steady-state quantile, where the CI approximately satisfies user-specified requirements on the CI's coverage probability and its absolute or relative precision. Sequest estimates a nonextreme quantile (i.e., its order is between 0.05 and 0.95) based on the methods of batching and sectioning. Sequem estimates extreme quantiles using a combination of batching, sectioning, and the maximum transformation. Two test problems show both the advantages and the limitations of these procedures. Based on the lessons learned in designing, justifying, implementing, and stress-testing Sequest and Sequem, we discuss future challenges in advancing the theory, algorithmic development, software implementation, performance evaluation, and practical application of improved procedures for steady-state quantile estimation.

1 INTRODUCTION

To evaluate long-run performance or risk for complex systems, steady-state simulations play a fundamental role in a wide range of application areas (Buzacott and Shanthikumar 1993; Conway 1963; Shortle et al. 2018; Trivedi 2002). On one hand, the steady-state expected value of a selected simulation output equals the long-run average of a time series of such outputs with probability one (Karlin and Taylor 1975, Theorem 5.5); on the other hand, a steady-state quantile of the selected output can gauge the long-run performance or risk associated with individual outputs (Glasserman 2004). For example, in a production-system simulation let X_i be the cycle time of the ith job (i.e., the job's time in the system), where $i \geq 1$. In the evaluation of an existing or proposed system design, an important performance measure may be $x_{0.95}$, the steady-state 0.95-quantile of the cycle-time distribution (Bekki et al. 2009, 2010) because as $i \to \infty$, the limiting probability is 95% that X_i does not exceed $x_{0.95}$.

To formalize our discussion, we assume that the simulation output process $\{X_i : i \geq 1\}$ is stationary with marginal cumulative distribution function (c.d.f.) $F(x) \equiv \Pr(X_i \leq x)$ and probability density function (p.d.f.) $f(x)$ for all $x \in \mathbb{R}$, where $f(x)$ is assumed to be continuous on its support. Given $p \in (0,1)$, the p-quantile of the marginal output distribution is $x_p \equiv F^{-1}(p) \equiv \inf\{x : F(x) \geq p\}$. If the time series $\{X_i : i = 1, \ldots, n\}$

consists of independent identically distributed (i.i.d.) outputs, then one can compute point and CI estimators of x_p based on a central limit theorem or nonparametric methods (Serfling 1980, §2.3.3 and §2.6.1).

If the simulation is not initialized in steady-state operation or the output process $\{X_i : i \geq 1\}$ is autocorrelated, then the estimation of steady-state quantiles involves substantial challenges. To address these challenges, several steady-state quantile estimation procedures been developed (Bekki et al. 2009, 2010; Chen and Kelton 2006, 2008; Drees 2003; Heidelberger and Lewis 1984; Iglehart 1976; Jain and Chlamtac 1985; McNeil and Frey 2000; Raatikainen 1987, 1990; Seila 1982a, 1982b). However, some existing or proposed estimation procedures for steady-state quantile estimation have significant drawbacks such as (a) lack of an adequate theoretical foundation; (b) implementation obstacles; (c) lack of effective guidelines for use in practical applications; (d) need for excessive user intervention; or (e) poor performance in practice, especially with respect to CI coverage probability, the size of the required sample, or the required computer resources. Issues (a)–(e) are elaborated in Alexopoulos et al. (2017, p. 22:3) and Alexopoulos et al. (2018b, §1).

In this paper we review two recent sequential procedures for estimating a steady-state quantile whose order p is given—namely, Sequest (Alexopoulos et al. 2018b), which is designed for estimating nonextreme quantiles (i.e., $0.05 \leq p \leq 0.95$); and Sequem (Alexopoulos et al. 2017), which is designed for estimating extreme quantiles. Section 2 of this paper provides a high-level overview of Sequest, and Section 3 provides a similar overview of Sequem. In Section 4 we summarize an experimental performance evaluation of these procedures in two queueing systems; and we discuss the insights gained and lessons learned in designing, justifying, implementing, and stress-testing these procedures. Finally in Section 5 we discuss future challenges in advancing the theory, algorithmic development, software implementation, performance evaluation, and practical application of improved procedures for steady-state quantile estimation.

2 OVERVIEW OF SEQUEST

Sequest exploits the methods of batching and sectioning to estimate x_p as follows. From the simulation-generated time series $\{X_1, \ldots, X_n\}$ of length $n = bm$, we form b nonoverlapping batches each of size m so that for $j = 1, \ldots, b$, the jth batch consists of the subseries $\{X_{(j-1)m+1}, \ldots, X_{jm}\}$. We sort the observations in the jth batch in ascending order to obtain the order statistics $X_{j,(1)} \leq X_{j,(2)} \leq \cdots \leq X_{j,(m)}$ and the conventional point estimator of x_p,

$$\check{x}_p(j,m) \equiv X_{j,(\lceil mp \rceil)}, \tag{1}$$

where $\lceil \cdot \rceil$ is the ceiling function. From the jth batch ($j = 1, \ldots, b$), both Sequest and Sequem compute the modified quantile estimator,

$$\widehat{x}_p(j,m) \equiv \begin{cases} X_{j,(1)} & \text{if } p \leq 0.5/m, \\ \delta_{p,m} X_{j,(\lceil mp+0.5 \rceil - 1)} + (1 - \delta_{p,m}) X_{j,(\lceil mp+0.5 \rceil)} & \text{if } 0.5/m < p < (m-0.5)/m, \\ X_{j,(m)} & \text{if } (m-0.5)/m \leq p, \end{cases} \tag{2}$$

where

$$\delta_{p,m} \equiv \lceil mp + 0.5 \rceil - (mp + 0.5) \tag{3}$$

(Avramidis and Wilson 1998, Equation (4)). Henceforth we refer to $\widehat{x}_p(j,m)$ as a batch quantile estimator (BQE). Similarly from the entire sample $\{X_1, \ldots, X_n\}$, we compute the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$ and the modified sectioning-based point estimator of x_p,

$$\widetilde{x}_p(n) \equiv \begin{cases} X_{(1)} & \text{if } p \leq 0.5/n, \\ \delta_{p,n} X_{(\lceil np+0.5 \rceil - 1)} + (1 - \delta_{p,n}) X_{(\lceil np+0.5 \rceil)} & \text{if } 0.5/n < p < (n-0.5)/n, \\ X_{(n)} & \text{if } (n-0.5)/n \leq p, \end{cases} \tag{4}$$

where $\delta_{p,n}$ is defined as in (3). Using Equations (2) and (4), we also compute a modified estimator of the variance of the BQEs,

$$\widetilde{S}^2_{\widehat{x}_p}(b,m) \equiv \frac{1}{b}\sum_{j=1}^{b}\left[\widehat{x}_p(j,m) - \widetilde{x}_p(n)\right]^2; \tag{5}$$

and based on Equations (4) and (5), we formulate the approximate $100(1-\alpha)\%$ CI for x_p,

$$\widetilde{x}_p(n) \pm t_{1-\alpha/2,\,b-1}\widetilde{S}_{\widehat{x}_p}(b,m)/b^{1/2}, \tag{6}$$

where $t_{u,\nu}$ is the u-quantile of Student's t-distribution with ν degrees of freedom for $u \in (0,1)$. Sequest adjusts the half-length of the CI (6) to compensate for the anomalous effects of correlation and skewness of the BQEs (2) that are used to compute the CI's half-length. See Alexopoulos et al. (2018b, §2) for a discussion of the theoretical, heuristic, and practical considerations leading to our use of the point estimator (4) and the CI estimator (6) as a starting point in the design of Sequest.

Sequest comprises four stages as summarized below.

Stage $\mathbb{S}_0$ initializes all the parameters of Sequest, including the batch size $m_0 = 128$, the batch count $b = 64$, several batch-size inflation factors, and the levels of significance used in testing the BQEs for approximate randomness.

Stage $\mathbb{S}_1$ consists of two steps:

- Step **[1]** encompasses two loops. The first loop starts with 64 batches of size 128 and iteratively increases the batch size by the inflation factor $\tau_{\text{wrm}} = 2$ until the BQEs in (2) have a sample standard deviation exceeding $\Delta_s = 10^{-10}$ and an estimated coefficient of variation exceeding $\Delta_c = 10^{-5}$. The objective of this loop is to determine a tentative batch size that is sufficiently large to contain a warm-up period in which a deterministic trend is the dominant effect; see Wang and Glynn (2016, §5.1) for an example of such a transient. Starting from the current batch size, the second loop iteratively applies the randomness test of von Neumann (1941) to the BQEs using a significance level (test size) α_{wrm} that decreases gradually from the initial value $\alpha_{\text{wrmi}} = 0.25$ to about $\alpha_{\text{wrmf}} = 0.001$. In particular, the size of the randomness test is iteratively updated using the assignment $\alpha_{\text{wrm}} \leftarrow \alpha_{\text{wrmi}}\left(0.6^{\ell-1}\right) + \alpha_{\text{wrmf}}\left(1-0.6^{\ell-1}\right)$, where ℓ is the iteration counter. Each time the BQEs fail the randomness test, the batch size is doubled, ℓ is incremented by 1, and the test is repeated.
- Step **[2]** executes a more powerful randomness test to determine a batch size that is large enough to ensure the BQEs are approximately i.i.d. and the sample skewness of the BQEs is approximately an unbiased estimator of the true skewness of the BQEs. This step starts by resetting the batch size to m_0 and by increasing the batch count to $b = 128$ so as to increase the power of the randomness test while keeping the batch size as small as possible. The level of significance α_{skw} of the randomness test decreases in the same way as in the second loop of step **[1]**, but the batch size increases by the smaller factor of $\tau_{\text{skw}} = 2^{1/2}$ after each failed test.

Stage $\mathbb{S}_2$ also consists of two steps:

- Step **[3]** starts by deleting the first w observations of the sample $\{X_1,\ldots,X_n\}$, where the data-truncation point w is the sum of the batch sizes from steps **[1]** and **[2]**. The truncated time series is reindexed as $\{Y_i = X_{w+i} : i = 1,\ldots,n^\dagger\}$, where $n^\dagger = n - w = bm$ and any needed extra observations are generated by resuming the simulation; then the BQEs $\{\widehat{y}_p(j,m)\}$ are updated using the $\{Y_i\}$. This step consists of a single loop that iteratively increases the batch size until the sample skewness $\widehat{\mathcal{B}}_{\widehat{y}_p}(b,m)$ of the updated BQEs has a magnitude below the threshold $\mathcal{B}^*(p) \equiv 0.65\exp\left(-0.5|p-0.5|^2\right)$. To avoid an explosion of the batch size while seeking to reduce the absolute skewness of the BQEs to a manageable level, we limit the number of iterations of step **[3]** to $u^* = 40$; and on each iteration

of step **[3]**, the batch size is updated according to the relation

$$m \leftarrow \left\lceil m \cdot \text{mid}\left\{ 1.05, \left[\widehat{\mathcal{B}}_{\widehat{y}_p}(b,m)/\mathcal{B}^*(p)\right]^2, \tau_{\text{skw}} \right\} \right\rceil , \tag{7}$$

where $\tau_{\text{skw}} = 2^{1/4}$ and $\text{mid}\{v_1, v_2, v_3\} \equiv v_{(2)}$.

- Step **[4]** increases the batch size by a factor of 4 and decreases the batch count b by a factor of 1/4 in an attempt to improve the coverage of the CI that will be computed in step **[6]**. This action neither increases the total sample size nor changes the bias of the sectioning-based point estimator of x_p. However, in general the absolute skewness and correlation of the recomputed BQEs decrease and the CI half-length increases owing to the increase in m and the decrease in b.

Stage $\mathbb{S}_3$ comprises three steps:

- Step **[5]** computes the warmed-up, sectioning-based point estimator $\widetilde{y}_p(n^\dagger)$ of x_p that is defined by the analogue of Equation (4) computed from the $\{Y_i\}$. This step also computes the correlation adjustment and the skewness adjustment that will be applied to the CI half-length H in step **[6]** so as to compensate for any remaining autocorrelation or skewness of the updated BQEs. Specifically, we compute the sample lag-one correlation $\widehat{\varphi}_{\widehat{y}_p}(b,m)$ of the updated BQEs and the associated correlation adjustment

$$A \leftarrow \max\left\{ \left[1 + \widehat{\varphi}_{\widehat{y}_p}(b,m)\right] / \left[1 - \widehat{\varphi}_{\widehat{y}_p}(b,m)\right], 1 \right\} .$$

  Moreover from the updated sample skewness $\widehat{\mathcal{B}}_{\widehat{y}_p}(b,m)$, we compute the associated skewness-adjustment parameter, $\beta \leftarrow \widehat{\mathcal{B}}_{\widehat{y}_p}(b,m) / \left(6\sqrt{b}\right)$; and we define the skewness-adjustment function,

$$G(\zeta) \equiv \begin{cases} \zeta & \text{if } |\beta| \le \varepsilon_s , \\ \dfrac{[1 + 6\beta(\zeta - \beta)]^{1/3} - 1}{2\beta} & \text{if } |\beta| > \varepsilon_s , \end{cases}$$

  where $\varepsilon_s = 10^{-3}$ and for all real u, we take $u^{1/3} \equiv \text{sign}(u)|u|^{1/3}$.
- Step **[6]** obtains Sequest's $100(1-\alpha)\%$ CI for x_p as follows. We compute the half-length of the correlation- and skewness-adjusted $100(1-\alpha)\%$ CI for the p-quantile x_p,

$$H \leftarrow \max\left\{ G(t_{1-\alpha/2,b-1}), G(t_{\alpha/2,b-1}) \right\} \left[A\widetilde{S}^2_{\widehat{y}_p}(b,m)/b \right]^{1/2} , \tag{8}$$

  to obtain the associated CI,

$$\widetilde{y}_p(n^\dagger) \pm H , \tag{9}$$

  which is adjusted for initialization bias as well as correlation and skewness of the BQEs.
- Step **[7]** executes the sequential run-length control logic as follows: if the half-length (8) of the current CI (9) satisfies the precision requirement $H \le H^*$, where

$$H^* = \begin{cases} r^* \left| \widetilde{y}_p(n^\dagger) \right| , & \text{for a user-specified relative precision level } r^* , \\ h^* , & \text{for a user-specified absolute precision level } h^* , \end{cases}$$

  then Sequest terminates; otherwise control returns to step **[5]** with the final batch count $b = 32$ and the updated batch size

$$m \leftarrow \left\lceil m \cdot \text{mid}\left\{ 1.02, (H/H^*)^2, 2 \right\} \right\rceil .$$

  The latter conservative assignment aims at curtailing any potential explosion of sample size and has its origins in recent sequential procedures for estimating the steady-state mean (Tafazzoli and Wilson 2011).

Sequest has been implemented in Java, and the package includes a graphical user interface that enables the user to do the following: (a) specify the parameters of any test process employed in Alexopoulos et al. (2017) or Alexopoulos et al. (2018b), and apply the procedure automatically to a sample path generated by the software in real time; or (b) apply the procedure semiautomatically to a dataset contained in a plain-text file. In either case, the user has the ability to specify an upper bound on the total sample size. If in case (b) the dataset is sufficiently large to allow normal termination of Sequest, then the algorithm delivers a CI for x_p that (approximately) satisfies the user-specified requirements on coverage and precision; otherwise, Sequest terminates with an estimate of the sample size required to continue execution in the current step.

We are currently developing a GitHub repository for three stand-alone versions of the Sequest package that can run under the Linux, Mac, and Windows operating systems, respectively. An online notice about the future availability of the software (Alexopoulos et al. 2018a) will be updated with a link to the GitHub repository upon its completion.

3 OVERVIEW OF SEQUEM

The Sequem procedure of Alexopoulos et al. (2017) combines the maximum transformation (Heidelberger and Lewis 1984) and the method of sectioning to convert the problem of estimating an extreme quantile x_p (so that $\max\{p, 1-p\} \in (0.95, 0.995]$) into the more tractable problem of estimating a nonextreme quantile of order $q \in [0.05, 0.95]$. We illustrate the maximum transformation when $p \in (0.95, 0.995]$. Suppose $V_1, \ldots, V_c$ are i.i.d. continuous random variables with c.d.f. $F_V(v)$, and we seek to estimate the p-quantile $v_p = F_V^{-1}(p)$. The random variable $Y \equiv \max\{V_1, \ldots, V_c\}$ has c.d.f. $F_Y(y) = \left[F_V(y)\right]^c$ for all y. It follows that $F_Y(v_p) = \left[F_V(v_p)\right]^c = p^c \equiv q$; hence $v_p = y_q = F_Y^{-1}(q)$, so estimating v_p is equivalent to estimating y_q. Most of the experimentation of Heidelberger and Lewis (1984, Tables I–IV) is based on setting c so that $p^c \approx 0.5$. However, this assignment of c often leads to excessive sample sizes when estimating extreme quantiles by a sequential procedure. For example, when $p = 0.99$ and $q = 0.5$, we have $c \approx \lfloor \ln(q)/\ln(p) \rfloor = 69$; and the impact of such a large value of c will become clear below. When $p \in [0.005, 0.05)$, one can apply the analogous minimum transformation detailed in Alexopoulos et al. (2017, §2).

Heidelberger and Lewis (1984) apply the maximum transformation to a dependent time series $\{X_i : i = 1, \ldots, n\}$ by (a) partitioning this dataset into L adjacent groups (subseries) each consisting of cm successive observations so that $n = cmL$; (b) organizing each subseries into c adjacent batches each of size m; and (c) conceptually arranging each subseries into a $c \times m$ matrix such that the first row of the matrix consists of the first batch, the second row consists of the second batch, etc. This arrangement is illustrated in Figure 1 for $c = 3$, $m = 3$, and $L = 3$. If the resulting L matrices of dimension $c \times m$ are concatenated horizontally (i.e., placed side by side) to form a $c \times (mL)$ matrix as in Figure 1, then each column of the $c \times (mL)$ matrix consists of observations separated by lag m; hence if m is sufficiently large, then the entries in each column are approximately i.i.d. with c.d.f. $F(x)$. For $\ell = 1, \ldots, L$ and $i = 1, \ldots, m$, we let $Y_{\ell,i}$ denote the maximum of the entries in column i within group ℓ; and we let $Y_{\ell,(1)} \le \cdots \le Y_{\ell,(m)}$ denote the associated order statistics within group ℓ. Finally from each group ℓ we compute the group quantile estimator (GQE) $Y_{\ell,(\lceil mq \rceil)}$ of y_q for $\ell = 1, \ldots, L$. If we take $c = \lfloor \ln(q)/\ln(p) \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function so that $q \approx p^c$, then for sufficiently large m the GQEs are approximately i.i.d. normal unbiased estimators of $y_q \approx x_p$; and an approximate CI for x_p can be computed from the sample mean and sample variance of the GQEs. Unfortunately, Heidelberger and Lewis (1984) do not provide a method for setting m and L; hence their nonsequential procedure requires substantial user intervention in practice, and their procedure is not readily extended to an automated sequential procedure for estimating extreme quantiles.

Sequem builds on the procedure of Heidelberger and Lewis (1984) in two respects:

- Sequem applies (a) the maximum transformation to estimate a p-quantile for $p \in (0.95, 0.995]$ using $q = 0.9$; and (b) the minimum transformation to estimate a p-quantile for $p \in [0.005, 0.05)$ using $q = 0.1$. These assignments for q yield substantially smaller values of c and thus substantially smaller values of the overall sample size n than are required when using $q = 0.5$.

	$\ell = 1$				$\ell = 2$				$\ell = 3$			
row 1	X_1	X_2	$\cdots$	X_m	X_{3m+1}	X_{3m+2}	$\cdots$	X_{4m}	X_{6m+1}	X_{6m+2}	$\cdots$	X_{7m}
row 2	X_{m+1}	X_{m+2}	$\cdots$	X_{2m}	X_{4m+1}	X_{4m+2}	$\cdots$	X_{5m}	X_{7m+1}	X_{7m+2}	$\cdots$	X_{8m}
row 3	X_{2m+1}	X_{2m+2}	$\cdots$	X_{3m}	X_{5m+1}	X_{5m+2}	$\cdots$	X_{6m}	X_{8m+1}	X_{8m+2}	$\cdots$	X_{9m}
max	$Y_{1,1}$	$Y_{1,2}$	$\cdots$	$Y_{1,m}$	$Y_{2,1}$	$Y_{2,2}$	$\cdots$	$Y_{2,m}$	$Y_{3,1}$	$Y_{3,2}$	$\cdots$	$Y_{3,m}$

Figure 1: Illustration of grouping mechanism with $c = 3$ and $L = 3$ groups.

- Sequem applies the sectioning method based on an alternative data layout in which the entire warmed-up (truncated) time series $\{X_i : i = 1, \ldots, n^\dagger = cmL\}$ is arranged conceptually into a $c \times (mL)$ matrix whose first row consists of the first subseries $\{X_i : i = 1, \ldots, mL\}$ of length mL, the second row consists of the second subseries $\{X_i : i = mL+1, \ldots, 2mL\}$ of length mL, etc. For $i = 1, \ldots, mL$, we let $\widetilde{Y}_i$ denote the maximum of the observations in column i of the $c \times (mL)$ matrix. Figure 2 depicts this arrangement when $c = 3$, $m = 3$, and $L = 3$. Such an arrangement of the dataset ensures that each column of the $c \times (mL)$ matrix consists of observations separated by lag $mL \gg m$ so those observations are nearly i.i.d. The sectioning-based point estimator of x_p is computed by sorting the $\left\{ \widetilde{Y}_i : i = 1, \cdots, mL \right\}$ in ascending order to obtain the order statistics $\widetilde{Y}_{(1)} \leq \cdots \leq \widetilde{Y}_{(mL)}$ so that we have the p-quantile estimator $\widetilde{y}_p(n^\dagger)$ computed from $\left\{ \widetilde{Y}_{(i)} : i = 1, \ldots, mL \right\}$ using an analogue of Equation (4) that is defined explicitly in Alexopoulos et al. (2017, Equation (15)).

We summarize briefly the four major stages composing Sequem. A detailed description and a flowchart of Sequem are given in Alexopoulos et al. (2017, §2).

Stage $\mathbb{S}_0$ initializes the parameters of Sequem, including $c = \lfloor \ln(q)/\ln(p) \rfloor$; the batch size $m_0 = 256$; the batch count $b = 64$; the sample size $n = bm = 16{,}384$; the group count $L = 64$; various batch-size inflation factors; and the sizes of the randomness tests that are applied to the BQEs used in Stage $\mathbb{S}_1$.

Stage $\mathbb{S}_1$ consists of two steps and does not employ the maximum transformation technique, working instead with the BQEs defined by Equation (2) as in Sequest. This is done mainly to avoid potentially excessive sample sizes in determining the length w of the warm-up period.

- Step **[1]** contains two loops that involve testing the BQEs for randomness. Paralleling the loops in step **[1]** of Sequest, these loops are designed to yield a batch size large enough to contain the main deterministic and stochastic transients that might affect the BQEs.
- Step **[2]** contains a loop that works similarly to its counterpart in Sequest by iteratively applying a higher-power randomness test to the BQEs, where the loop's starting batch size is $m = m_0$ and the batch count is reset to $b = \min\{cL, 256\}$. This loop is designed to yield a batch size large enough to contain any remaining transients that might affect the BQEs or GQEs.

Stage $\mathbb{S}_2$ contains two steps:

- Step **[3]** starts by deleting the first w observations in the original dataset, where w is the sum of the batch sizes determined in steps **[1]** and **[2]**. We initialize the maximum transformation method by collecting enough additional data to form L groups of size cm, starting with the batch size $m = m_0 = 256$. This grouping is retained in steps **[3]**–**[5]**. Henceforth for simplicity we let $\{X_i : i = 1, \ldots, n^\dagger = cmL\}$ denote the warmed-up (truncated) dataset. The remainder of this step consists of a single loop that iteratively increases the batch size until the estimated absolute skewness of the GQEs falls below the fixed threshold $\mathcal{B}^*(p) \equiv 0.60$. To avoid excessive batch sizes, this loop

	$\ell=1$				$\ell=2$				$\ell=3$			
row 1	X_1	X_2	$\cdots$	X_m	X_{m+1}	X_{m+2}	$\cdots$	X_{2m}	X_{2m+1}	X_{2m+2}	$\cdots$	X_{3m}
row 2	X_{3m+1}	X_{3m+2}	$\cdots$	X_{4m}	X_{4m+1}	X_{4m+2}	$\cdots$	X_{5m}	X_{5m+1}	X_{5m+2}	$\cdots$	X_{6m}
row 3	X_{6m+1}	X_{6m+2}	$\cdots$	X_{7m}	X_{7m+1}	X_{7m+2}	$\cdots$	X_{8m}	X_{8m+1}	X_{8m+2}	$\cdots$	X_{9m}
max	$\widetilde{Y}_1$	$\widetilde{Y}_2$	$\cdots$	$\widetilde{Y}_m$	$\widetilde{Y}_{m+1}$	$\widetilde{Y}_{m+2}$	$\cdots$	$\widetilde{Y}_{2m}$	$\widetilde{Y}_{2m+1}$	$\widetilde{Y}_{2m+2}$	$\cdots$	$\widetilde{Y}_{3m}$

Figure 2: Illustration of the sectioning mechanism for the derivation of the final point estimator of x_p with $c=3$ and $L=3$ groups.

is limited to 50 iterations, has a temporary upper bound $n^* = 30\times 10^8$ on the total sample size, and increases the batch size by a factor similar to Equation (7) on successive iterations of the loop.

- Step **[4]** functions similarly to the corresponding step of Sequest by halving the number of groups L and doubling the batch size m.

Stage $\mathbb{S}_3$ includes three steps:

- Step **[5]** computes the CI half-length adjustments designed to compensate for any remaining autocorrelation and skewness of the GQEs as in the corresponding step of Sequest.
- Step **[6]** involves (a) conceptually rearranging the truncated dataset as shown in Figure 2 to compute the sectioning-based point estimator of x_p; and (b) computing the CI estimator $\widetilde{y}_p(n^\dagger)\pm H$ of x_p whose half-length H incorporates the correlation and skewness adjustments from step **[5]**.
- Step **[7]** is a direct analogue of step **[7]** of Sequest that executes the sequential run-length control logic.

4 EXPERIMENTAL RESULTS

This section contains experimental results for two test processes. The first test process consists of the total queue-waiting time (prior to service) for customers in a system comprised of two M/M/1 queues in tandem. In this case we consider values of $p\in\{0.30,0.50,0.70,0.90,0.95,0.995\}$. The second test process consists of times in system for an M/G/1 queue with service times that are a mixture of two gamma distributions, In this process we augment the set of values of p to $\{0.05,0.10,0.15,0.20,0.25,0.30,0.50,0.70,0.90,0.95,0.995\}$ in order to study the potential effect of the bimodality of the marginal p.d.f. on the variability of the sample size near the two modes. For each problem we considered two levels of relative precision: $r^*=\infty$ (no precision requirement) and $r^*=0.02$; the latter value was chosen to evaluate the effectiveness of the sequential mechanism on step **[7]** of both Sequest and Sequem when relatively little additional sampling is required beyond the sample size required under no precision requirement. All experiments were performed on a Windows 7 desktop computer equipped with an Intel Core i7-3770 CPU and 16 GB of RAM.

All the experimental results tabulated below are based on 1000 independent replications of each test process. In each table the entries in bold represent the relevant performance measures for Sequem. In each table, columns 2 through 4 respectively list the following: the nearly exact value of each quantile under study; the corresponding average value of the final point estimator $\widetilde{y}_p(n^\dagger)$ of x_p averaged over all 1000 replications, where $n^\dagger = n-w$ is the final truncated sample size; and Avg. $\left|\widetilde{y}_p(n^\dagger)-x_p\right|$ denotes the average of the point estimator's absolute bias computed over all 1000 replications. For nominal 95% CIs, columns 5 and 6 respectively contain the average half-length $\overline{H}$ and average relative precision, where the latter is the average CI half-length expressed as a percentage of the magnitude of the average CI midpoint, that is, $100\times\overline{H}/\left|\text{Avg. }\widetilde{y}_p(n^\dagger)\right|$%; and column 7 contains the estimated CI coverage probability. The last two columns list the respective final batch size and sample size. The disparity between the batch sizes used

by Sequest and Sequem reflects the impact of the parameter $c = \lfloor \ln(0.9)/\ln(p) \rfloor$. Since each data group consists of cm consecutive observations, this disparity grows as p increases.

4.1 An M/M/1/M/1 Queue-Waiting-Time Process

The first test process consists of the total queue-waiting time for a customer in a tandem system composed of two M/M/1 queues. This system has arrival rate $\lambda = 1$ at the first queue and service rate $\omega = 1.25$ at each server. This system was initialized in the empty-and-idle state. In steady-state each server has a utilization of $\rho = 0.8$, and the expected value of a customer's total queue-waiting time is equal to 6.4. The steady-state c.d.f. $F(x)$ of a customer's total waiting time is

$$F(x) = \frac{[\lambda + 2(1-\rho)\omega + \lambda\omega(1-\rho)x]e^{-\omega(1-\rho)x} - 2\omega(1-\rho)^2 - 2\lambda(1-\rho) - \lambda}{\lambda - 2\omega} \quad \text{for} \quad x \geq 0$$

(Karpelevitch and Kreinen 1992, Theorem 2). We computed the nearly exact value of x_p for each selected value of p by inverting a piecewise-linear approximation to $F(x)$ for $x \in [0, 75]$ based on increments of size $\Delta x = 10^{-3}$.

Table 1 summarizes the experimental results for this test process. The results for $p \geq 0.95$ appeared in Alexopoulos et al. (2017, §3.2). With no precision requirement and $p \in [0.3, 0.95]$, Sequest's performance was good because the estimated CI coverage probabilities were close to the nominal 95% level, the point estimates $\widetilde{y}_p(n^\dagger)$ exhibited low average absolute bias (ranging between 0.041 and 0.183), and the average sample sizes were judged to be reasonable. Unfortunately, Sequest's performance deteriorated substantially for $p > 0.95$: the estimated CI coverage probabilities dropped from the near-nominal value of 94.7% for $p = 0.95$ to the substandard value of 87.6% for $p = 0.99$ and the unacceptable value of 81.1% for $p = 0.995$. On the other hand, Sequem's performance was consistently good for $p \geq 0.95$. With a relative precision level of 2%, both procedures performed equally well, requiring roughly equal average sample sizes to deliver CIs whose average levels of coverage and relative precision conformed closely to their respective nominal levels.

4.2 An M/G/1 Time-in-System Process

The second example involves an M/G/1 queueing system with mean interarrival time 10 and i.i.d. service times whose distribution is a mixture of two gamma distributions: (a) with probability 0.8, the service time has a three-parameter gamma distribution with location parameter $\gamma_1 = 2$, shape parameter $\alpha_1 = 2$, and scale parameter $\beta_1 = 2$ so the associated p.d.f. has mode 4 and mean 6; and (b) with probability 0.2 the service time has a three-parameter gamma distribution with location parameter $\gamma_2 = 12$, shape parameter $\alpha_2 = 4$, and scale parameter $\beta_2 = 1$ so the associated p.d.f. has mode 15 and mean 16. Since the mean service time is equal to $0.8(6) + 0.2(16) = 8$, the traffic intensity is $\rho = 0.8$. The simulation-generated response of interest X_i ($i \geq 1$) is the total time in the system (including service time) for the ith departing customer. The Pollaczek-Khinchin formula yields the steady-state mean time in the system of 29.8. This system was initialized in the empty-and-idle state.

Figure 3 depicts a histogram of 10^8 observations of time in the system after deleting the first 100,000 observations. Figure 3 leads to the conclusion that the steady-state p.d.f. $f(x)$ of the $\{X_i\}$ has two modes near 5 and 15 and an antimode near 11.

Table 2 summarizes the performance of Sequest and Sequem for this M/G/1 queueing system. As in the M/M/1/M/1 queueing system, we see that with no precision requirement Sequest's performance is good for $p \in [0.05, 0.95]$; but there is noticeable degradation in CI coverage for the extreme quantile $x_{0.995}$. On the other hand, with no precision requirement Sequem outperforms Sequest when $p \in [0.95, 0.995]$, requiring significantly smaller sample sizes than Sequest while delivering CIs that consistently exhibit close conformance to the nominal coverage probability. Finally with the relative precision requirement $r^* = 0.02$, both Sequest and Sequem exhibited comparably good performance.

An examination of the entries of Table 2 also reveals a noticeable variation in the average sample sizes required to estimate quantiles of low order ($0.05 \leq p \leq 0.30$) with no precision requirement. The following

Table 1: Performance of Sequest- and Sequem-delivered point and 95% CI estimators of the p-quantile x_p of the M/M/1/M/1 queue-waiting-time process described in Section 4.1 based on 1000 replications. The outcomes from Sequem are in bold typeface.

No CI Precision Requirement								
p	x_p	Avg. $\widetilde{y}_p(n^\dagger)$	Avg. $\lvert\text{Bias}[\widetilde{y}_p(n^\dagger)]\rvert$	$\overline{H}$	Avg. CI Rel. Prec. (%)	CI Cover. (%)	$\overline{m}$	$\overline{n}$
0.300	2.748	2.746	0.041	0.111	4.057	96.4	10,571	338,750
0.500	5.079	5.078	0.063	0.172	3.391	96.5	11,505	368,704
0.700	8.126	8.122	0.088	0.240	2.961	96.4	15,734	504,095
0.900	13.931	13.912	0.138	0.347	2.493	95.4	35,988	1,152,326
0.950	17.349	17.313	0.183	0.430	2.481	94.7	59,728	1,912,034
		17.313	**0.304**	**0.833**	**4.812**	**95.0**	**10,176**	**652,442**
0.990	24.928	24.589	0.501	0.703	2.858	87.6	201,486	6,448,335
		24.889	**0.324**	**0.816**	**3.280**	**94.9**	**11,561**	**3,701,075**
0.995	28.096	27.499	0.771	0.864	3.142	81.1	329,066	10,530,896
		28.080	**0.329**	**0.834**	**2.969**	**94.8**	**10,748**	**7,224,510**
CI Relative Precision = 2%								
p	x_p	Avg. $\widetilde{y}_p(n^\dagger)$	Avg. $\lvert\text{Bias}[\widetilde{y}_p(n^\dagger)]\rvert$	$\overline{H}$	Avg. CI Rel. Prec. (%)	CI Cover. (%)	$\overline{m}$	$\overline{n}$
0.300	2.748	2.747	0.019	0.049	1.793	96.7	38,778	1,241,390
0.500	5.079	5.076	0.034	0.091	1.787	95.2	30,275	969,348
0.700	8.126	8.124	0.053	0.144	1.771	96.0	31,373	1,004,546
0.900	13.931	13.924	0.088	0.237	1.703	96.2	50,537	1,617,912
0.950	17.349	17.343	0.108	0.286	1.649	96.5	76,926	2,462,381
		17.343	**0.123**	**0.312**	**1.799**	**95.1**	**33,705**	**2,158,319**
0.990	24.928	24.916	0.142	0.374	1.503	96.0	250,911	8,029,952
		24.927	**0.176**	**0.436**	**1.747**	**94.3**	**23,411**	**7,493,148**
0.995	28.096	28.081	0.155	0.413	1.472	94.8	414,067	13,250,919
		28.099	**0.191**	**0.482**	**1.717**	**94.3**	**18,943**	**12,731,353**

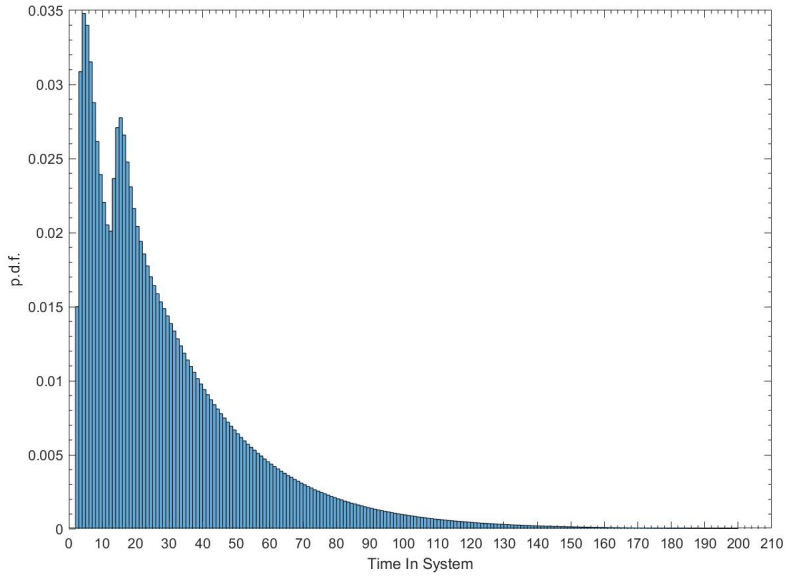

Figure 3: Histogram of 100M response times from the M/G/1 model.

discussion aims at a plausible explanation of this phenomenon. In the absence of a precision requirement, the final value of the batch size m is determined mainly by step **[3]** of Sequest, wherein m is iteratively increased based on Equation (7) until the sample skewness $\widehat{\mathcal{B}}_{\widehat{y}_p}(b,m)$ of the BQEs $\left\{\widehat{y}_p(j,m) : j = 1,\ldots,b\right\}$

Table 2: Performance of Sequest- and Sequem-delivered point and 95% CI estimators of the p-quantile x_p of the M/G/1 time-in-system process described in Section 4.2 based on 1000 replications. The outcomes from Sequem are in bold typeface.

No CI Precision Requirement								
p	x_p	Avg. $\widetilde{y}_p(n^\dagger)$	Avg. $\lvert\text{Bias}[\widetilde{y}_p(n^\dagger)]\rvert$	$\overline{H}$	Avg. CI Rel. Prec. (%)	CI Cover. (%)	$\bar{m}$	$\bar{n}$
0.005	4.121	4.122	0.021	0.056	1.430	96.1	5,430	174,061
0.100	5.564	5.563	0.029	0.085	1.527	97.1	8,324	266,702
0.150	7.128	7.130	0.040	0.120	1.689	96.9	10,808	346,192
0.200	8.957	8.961	0.061	0.176	1.960	96.3	11,626	372,391
0.250	11.141	11.150	0.141	0.399	3.577	96.8	4,519	144,944
0.300	13.547	13.538	0.152	0.420	3.100	95.7	3,684	118,250
0.500	21.894	21.906	0.170	0.484	2.210	96.8	10,288	329,598
0.700	35.312	35.320	0.303	0.829	2.346	96.2	12,871	412,293
0.800	45.905	45.937	0.404	1.058	2.303	95.2	16,522	529,148
0.900	64.000	64.032	0.506	1.367	2.135	96.7	28,147	901,183
0.950	82.069	82.152	0.629	1.657	2.016	95.7	48,665	1,557,796
		82.095	**1.275**	**3.680**	**4.483**	**96.3**	**7,920**	**507,779**
0.990	124.009	123.478	1.459	2.708	2.193	93.4	176,894	5,661,127
		124.140	**1.531**	**3.997**	**3.220**	**94.5**	**8,744**	**2,808,851**
0.995	142.088	140.703	2.414	3.308	2.351	89.5	308,685	9,787,453
		142.274	**1.659**	**4.204**	**2.955**	**95.1**	**8,299**	**5,578,501**
CI Relative Precision = 2%								
p	x_p	Avg. $\widetilde{y}_p(n^\dagger)$	Avg. $\lvert\text{Bias}[\widetilde{y}_p(n^\dagger)]\rvert$	$\overline{H}$	Avg. CI Rel. Prec. (%)	CI Cover. (%)	$\bar{m}$	$\bar{n}$
0.050	4.121	4.122	0.021	0.056	1.361	95.5	5,544	177,707
0.100	5.564	5.563	0.029	0.080	1.437	97.1	8,607	275,741
0.150	7.128	7.131	0.039	0.108	1.517	96.4	11,576	370,766
0.200	8.957	8.963	0.055	0.145	1.622	96.0	13,603	435,641
0.250	11.141	11.147	0.076	0.197	1.768	95.5	13,692	438,505
0.300	13.547	13.553	0.094	0.244	1.801	95.7	9,064	290,394
0.500	21.894	21.902	0.144	0.374	1.707	96.1	13,629	436,517
0.700	35.312	35.327	0.235	0.606	1.715	95.8	18,273	585,152
0.800	45.905	45.922	0.306	0.784	1.708	96.0	22,965	735,337
0.900	64.000	64.034	0.409	1.068	1.669	96.2	35,474	1,135,669
0.950	82.069	82.149	0.521	1.328	1.617	95.4	57,691	1,846,626
		82.173	**0.581**	**1.476**	**1.796**	**94.9**	**24,263**	**1,553,720**
0.990	124.009	124.283	0.715	1.833	1.475	95.7	198,070	6,338,767
		124.283	**0.884**	**2.180**	**1.754**	**94.0**	**17,334**	**5,547,929**
0.995	142.088	142.308	0.775	1.993	1.401	95.7	349,806	11,194,297
		142.361	**1.005**	**2.456**	**1.725**	**95.1**	**14,215**	**9,553,620**

falls below the threshold $\mathcal{B}^*(p)$ in magnitude. Let $\text{Sk}\big[\widehat{y}_p(1,m)\big]$ denote the true skewness of $\widehat{y}_p(1,m)$. Here we analyze heuristically the dependence of $\text{Sk}\big[\widehat{y}_p(1,m)\big]$ on the value of p for a given process $\{X_i : i \geq 1\}$. Based on an analogous result derived in Alexopoulos et al. 2018b for the case which the $\{X_i\}$ are i.i.d., we postulate the following asymptotic relation:

$$\text{Sk}\big[\widehat{y}_p(1,m)\big] \sim \frac{\mathfrak{N}(p)}{[\mathfrak{D}(p)]^{3/2} m^{1/2}} \left\{ \frac{2(1-2p)}{[p(1-p)]^{1/2}} - \frac{3f'(x_p)[p(1-p)]^{1/2}}{[f(x_p)]^2} \right\} \text{ as } m \to \infty, \tag{10}$$

where the functions $\mathfrak{N}(p)$ and $\mathfrak{D}(p)$ are ultimately determined by the stochastic dependency structure of the $\{X_i : i \geq 1\}$. Equation (10) provides some heuristic insight into how $\text{Sk}\big[\widehat{y}_p(1,m)\big]$ is determined by p, $f(x_p)$, $f'(x_p)$, and the stochastic dependency structure of the process $\{X_i\}$. We concluded that in general when Sequest is applied without a precision requirement, there could be significant variation in the required sample size and in the resulting relative or absolute precision of the delivered CI as those statistics depend on the selected values of p. We also believe that even when Sequest is applied with a precision requirement, the foregoing analysis can provide some insight into the variation of the sample sizes required by Sequest across different values of p.

5 CONCLUSIONS

In this article we reviewed the recent Sequest and Sequem sequential procedures designed for estimating nonextreme and extreme steady-state quantiles x_p, respectively. We also used two numerical examples to evaluate the performance of these methods over an extended range of values of the order p. The experimental results were aligned with the extensive experimentation in Alexopoulos et al. (2017, 2018b): both methods yielded point and CI estimates with near-nominal coverage probabilities within their domain of applicability with substantially lower average sample sizes than existing methods. Our future work on steady-state quantile estimation will focus on (i) development of a unified method for delivering point and CI estimators of x_p having uniformly good performance for all $p \in [0.005, 0.995]$; and (ii) implementation of the method in robust public-domain software for both real-time and off-line use.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grants CMMI-1232998/1233141.

REFERENCES

Alexopoulos, C., D. Goldsman, A. C. Mokashi, and J. R. Wilson. 2017. "Automated Estimation of Extreme Steady-State Quantiles via the Maximum Transformation". *ACM Transactions on Modeling and Computer Simulation* 27(4):22:1–22:29.

Alexopoulos, C., D. Goldsman, A. C. Mokashi, K.-W. Tien, and J. R. Wilson. 2018a. "Notice on the Future Availability of the Sequest Software in a GitHub Repository". Accessed July August 5, 2018. http://www4.ncsu.edu/~jwilson/files/sequest-availability.pdf.

Alexopoulos, C., D. Goldsman, A. C. Mokashi, K.-W. Tien, and J. R. Wilson. 2018b. "Sequest: A Sequential Procedure for Estimating Quantiles in Steady-State Simulations". *Operations Research* In review. Accessed August 5, 2018. http://www4.ncsu.edu/~jwilson/files/sequest18or.pdf.

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 M. Kulahci. 2009. "Simulation-Based Cycle-Time Quantile Estimation in Manufacturing Settings Employing Non-FIFO Dispatching Policies". *Journal of Simulation* 3:69–83.

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.

Buzacott, J. A., and J. G. Shanthikumar. 1993. *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, New Jersey: Prentice-Hall.

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 & Operations Research* 35(4):1003–1016.

Conway, R. W. 1963. "Some Tactical Problems in Digital Simulation". *Management Science* 10(1):47–61.

Drees, H. 2003. "Extreme Quantile Estimation for Dependent Data, with Applications to Finance". *Bernoulli* 9(1):617–657.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. New York: Springer-Verlag.

Heidelberger, P., and P. A. W. Lewis. 1984. "Quantile Estimation in Dependent Sequences". *Operations Research* 32(1):185–209.

Iglehart, D. L. 1976. "Simulating Stable Stochastic Systems, VI: Quantile Estimation". *Journal of the Association for Computing Machinery* 23(2):347–360.

Jain, R., and I. Chlamtac. 1985. "The P^2 Algorithm for Dynamic Calculation of Quantiles and Histograms without Storing Observations". *Communications of the ACM* 28(10):1076–1085.

Karlin, S., and H. M. Taylor. 1975. *A First Course in Stochastic Processes*. 2nd ed. New York: Academic Press.

Karpelevitch, F. I., and A. Y. Kreinen. 1992. "Joint Distributions in Poissonian Tandem Queues". *Queueing Systems* 12:273–286.

McNeil, A. J., and R. Frey. 2000. "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach". *Journal of Empirical Finance* 7:271–300.

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.

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* 39(6):193–200.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

Shortle, J. F., J. M. Thompson, D. Gross, and C. M. Harris. 2018. *Fundamentals of Queueing Theory*. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

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.

Trivedi, K. S. 2002. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*. 2nd ed. New York: John Wiley & Sons.

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.

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):1:1–1:30.

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 an active participant in the Winter Simulation Conference, having been *Proceedings* Co-Editor in 1995, Associate Program Chair in 2006, and a member of the Board of Directors of WSC during the period 2008–2016. 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.

DAVID GOLDSMAN is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests include simulation output analysis, ranking and selection, and healthcare simulation. He was Program Chair of the Winter Simulation Conference in 1995, and a member of the WSC Board of Directors between 2001–2009. He was also a trustee of the WSC Foundation. His e-mail address is sman@gatech.edu, and his Web page is www.isye.gatech.edu/∼sman.

ANUP C. MOKASHI is an operations research development tester for SAS Simulation Studio at the SAS Institute. He is a member of IISE and INFORMS. His e-mail address is Anup.Mokashi@sas.com.

JAMES R. WILSON is an emeritus professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. His research interests concern the design and analysis of simulations used in healthcare and production engineering. As a WSC participant, he served as *Proceedings* Editor (1986), Associate Program Chair (1991), and Program Chair (1992). During the period 1997–2004, he was a member of the WSC Board of Directors. He is a member of ACM and ASA, and he is a Fellow of IISE and INFORMS. His e-mail address is jwilson@ncsu.edu, and his Web page is www.ise.ncsu.edu/jwilson.