Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

ESTIMATING STEADY STATE MEAN FROM SHORT TRANSIENT SIMULATIONS

Pieter A. Voss
Jorge Haddock
Thomas R. Willemain

Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute
Troy, New York 12180-3590, USA

ABSTRACT

This paper presents a new approach to estimating and constructing confidence intervals for the steady state mean of a stochastic process from short simulations which may exhibit significant transient response. Specifically, we examine the conditional least squares estimator for the mean of an autoregressive process. If the process is autoregressive with normal innovations, this estimator is the conditional maximum likelihood estimate (MLE) of the mean. We show that the MLE is asymptotically normal, and derive a finite-sample approximation to its distribution. This provides the basis for two asymptotically valid single-replication confidence intervals which do not require choosing a batch size. As a point estimator, the MLE is a generalization of the estimator due to Snell and Schruben (1984) and is related to the weighted batch mean (Bischak, Kelton, and Pollock, 1993).

Empirical results for a queuing network show that the autoregressive process is a reasonable model of transient response. For short series with reasonable initializations (e.g., empty and idle), the MLE yields confidence intervals which are comparable or superior to those of existing procedures, in both single and parallel replication simulations.

1 INTRODUCTION

In this paper we present a new approach to the problem of estimating and constructing confidence intervals for the steady state mean μ of a stochastic process from short simulations which may exhibit significant transient response. We assume that the process may be strongly autocorrelated (positive or otherwise), that sufficient knowledge of steady state conditions is not available to initialize the process in steady state, and that the cost of collecting simulation observations is large so that available data is limited to a finite series of n observations $\{X_1, \ldots, X_n\}$.

The problem of estimating the steady state mean in general, and of controlling the effects of initial transient response in particular have been widely studied; many existing methods are surveyed by Law and Kelton (1991) and Pawlikowski (1990), and progress continues to be made. However, all existing confidence interval procedures explicitly assume that the observed process is in steady state, stationary in both mean and covariance. These include (unweighted) batch means (see Law, 1977), Fishman's autoregressive approach (Fishman, 1978), standardized time series (Schruben, 1983), weighted batch means (Bischak, Kelton, and Pollock, 1993), and intervals based on the ratio estimator due to Glynn and Heidelberger (1992). Since initial transient effects may invalidate the assumption of stationarity, either truncation or initialization methods are employed to minimize bias due to initial transient response. A third approach, weighting initial observations, is due to Snell and Schruben (1984) and is developed further in this paper.

For short transient simulations, there is as yet no completely satisfactory confidence interval procedure. In particular, Sargent, Goldsman, and Swain (1992) show that most confidence interval procedures are invalid for small sample sizes, even when assumptions such as stationarity and normality are satisfied. Constructing valid confidence intervals from short stationary series is the motivation for weighted batch means (Bischak, Kelton and Pollock, 1993). However, their derivation assumes the process is in steady state.

The ability to analyze short series is particularly valuable when simulating for long periods of time is costly. When the cost of simulating is high, parallelization is a simple and direct method of exploiting parallel processors to obtain results in shorter "wall-clock" time. In this context, two ways in which the problem of confidence interval estimation can arise are:

(1) Different simulation models may be independently executed on separate processors; confidence intervals may be periodically computed for each model and particular systems may be then selected or rejected on the basis of the results.

(2) Independent replications of the same simulation model may be executed on parallel processors, and the results of each simulation run are then combined to form a single estimate and confidence interval.

While the above situations lend themselves readily to parallel simulation, similar situations also occur when simulating on single-processor computers.

This paper investigates the conditional maximum likelihood estimator for the mean of autoregressive processes. Based on two expressions for its asymptotic distribution, we derive asymptotically valid confidence intervals procedures for both single-replication and multiple-replication experiments. These confidence intervals are often superior to those obtained by existing methods in short series with significant initialization bias. In long series, the estimator reduces to the sample mean, and the confidence intervals are asymptotically equivalent to those of existing procedures. Thus the new procedure applies to simulations of any length.

2 AUTOREGRESSIVE MODEL

The presence of transient behavior necessarily indicates that the process is autocorrelated, that is, sequential observations are related to previous observations. The principal distinction between transient and steady state observations is that initial observation are typically more extreme than steady state observations, and they tend to be systematically so (e.g., initializing a queuing simulation empty and idle). To study the transient response, we therefore require a model of the process which captures this dependence. While no single model is able to capture every situation, the linear autoregressive process is widely used to approximate the autocorrelation structure of a wide range of discrete-time stochastic processes. Further, autoregressive processes exhibit the sum-of-exponentials transient mean characteristic of many short-memory stochastic processes. Processes for which transient response has been theoretically shown to be bounded by sum-of-exponential functions include ARMA processes (Fuller, 1976), continuous time Markov chains (Glynn, 1984) and M/M/k queuing systems (Odoni and Roth, 1983).

The autoregressive model assumes that each new observation X_t is linearly related to previous observations with the addition of a random "shock" or innovation. An autoregressive process of order p, AR(p), is represented as follows:

$$X_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t, \; t = 1, 2, \ldots \qquad (1)$$

where p is finite and $\{\varepsilon_t, t = 1, 2, \ldots\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance $\sigma_\varepsilon^2 < \infty$. The value $\phi = \{\phi_0, \ldots, \phi_p\}$ is a vector of autoregressive coefficients. Provided tnat the mean system response converges to a unique limiting steady state, the steady state mean response $\mu = \mu(\phi)$ is:

$$\mu(\phi) = \lim_{t \to \infty} E[X_t] = \phi_0 \left(1 - \sum_{i=1}^{p} \phi_i \right)^{-1}. \qquad (2)$$

Although there may exist a value for (2), it is not necessarily true that mean system response will converge to a unique finite value. For example, consider a queue where customers arrive slightly more rapidly than they can be served, in which case the sequence of time-in-system for each customer will diverge. While this will become apparent after many observations, it may not be immediately apparent from the first customers that the server cannot keep up with demand. If it is not known in advance that the server can keep up with demand, it may be necessary to empirically establish the existence of a steady state mean before estimating its value.

Taking the expectation of (1) yields the linear difference equation $EX_t - \sum \phi_i EX_{t-i} = 0$. Solving this equation yields the sequence of expected values, called the transient expectation function:

$$EX_t = \frac{\phi_0}{1 - \sum_{i=1}^{p} \phi_i} + \sum_{i=1}^{p'} \sum_{j=1}^{n_i} c_{ij} t^j r_i^{-t}. \qquad (3)$$

where c_{ij} are constants determined by the initial p observations, and $\{r_1, \ldots, r_p\}$ are the roots to the characteristic equation $1 - \sum_{i=1}^{p} \phi_i x^i = 0$. Since the roots are not necessarily distinct, we let n_i denote the multiplicity of root r_i. From (3), we can see that the transient response decays exponentially to zero for large t if and only if $|r_i| > 1, i = 1, \ldots, p$; otherwise, the transient response grows exponentially. This stationarity condition can also be written as:

$$1 - \sum_{i=1}^{p} \phi_i x^i = 0 \quad \forall |x| \le 1 \qquad (4)$$

3 ESTIMATION OF STEADY STATE MEAN

Since we assume initialization effects may be large, we must estimate the autoregressive model parameters without assuming that the process is stationary. We choose here the method of maximum likelihood, conditioning on the first p observations. For the purpose of deriving the

maximum likelihood estimator and its distribution, the order of autoregression p is assumed known. In practice, it is also necessary to identify the order of autoregression; however, to establish consistency of the estimator, it is sufficient to assume that the estimators of the autoregressive order and coefficients are consistent (Voss, Haddock, and Willemain, 1996).

Let $\hat{\phi}$ be an estimate of the vector of autoregressive coefficients ϕ. We can define an estimate for the mean $\hat{\mu} = \mu(\hat{\phi})$ as follows:

$$\mu(\hat{\phi}) = \begin{cases} \hat{\phi}_0 \left(1 - \sum_{j=1}^{p} \hat{\phi}_j \right)^{-1}, & \text{if } 1 - \sum_{i=1}^{p} \hat{\phi}_i x^i \neq 0 \quad \forall |x| \leq 1 \\ \text{undefined}, & \text{otherwise} \end{cases} \tag{5}$$

Note that this estimator does not require the assumption that a unique limiting steady state response exists. If the estimated autoregressive parameters satisfy the stationarity condition, then the estimate is defined and finite. On the other hand, if the estimated parameters fall outside the stationary region, system response is to all indications nonstationary in mean or covariance, and no estimate of steady state is meaningful. If $\hat{\phi}$ is a consistent estimator of ϕ, and $\{X_t\}$ converges to a covariance stationary process, then $\hat{\mu}$ is defined almost surely and is a consistent estimator of μ (Voss, Haddock, and Willemain, 1996).

We can compute the conditional least squares estimate of ϕ as follows (Fuller, 1976):

$$\hat{\phi} = A_n^{-1} v_n \tag{6}$$

where

$$A_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \underline{X}_t' \underline{X}_t$$

$$\underline{X}_t = (1, X_{t-1}, \ldots, X_{t-p})$$

$$v_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \underline{X}_t' X_t,$$

Provided that the stationarity condition (4) is satisfied, this estimator is asymptotically normally distributed:

$$n^{1/2} (\hat{\phi} - \phi) \rightarrow N\left(0, A^{-1} \sigma_\varepsilon^{\,2} \right), \tag{7}$$

where $A = p \lim_{n \to \infty} A_n$. The elements of A are given by

$$a_{ij} = \begin{cases} \gamma_X(|i-j|) + \mu^2; & i,j=2,3, \ldots, p+1 \\ 1; & i=j=1 \\ \mu; & \text{otherwise} \end{cases}, \tag{8}$$

where $\gamma_X(k)$ is the process autocovariance at lag k. The asymptotic properties for the linear least squares estimators also apply to estimates obtained by any method which is asymptotically equivalent to the method of maximum likelihood, such as solution of the Yule-Walker equations, which differ from the linear least squares estimates by a quantity which approaches zero in probability as n increases (Fuller, 1976).

When model (1) holds and the innovations ε_t are normally distributed, then $\hat{\phi}$ is the conditional maximum likelihood estimate of ϕ. Further, wherever $\mu(\hat{\phi})$ is defined, it is the maximum likelihood estimate of μ. This follows from the invariance principle, since the mapping $\mu: \Re^{p+1} \rightarrow \Re$ defines a continuous interval on $\Re$, and $\mu(\hat{\phi})$ is locally differentiable with respect to ϕ. For this reason, we shall refer to $\hat{\mu}$ as the MLE.

Voss, Haddock and Willemain (1996) show that the estimator $\hat{\mu}$ is related to other estimators in the literature. In particular, $\hat{\mu}$ is equivalent to the weighted batch mean Bischak, Kelton, and Pollock (1993) when all the data are grouped into a single batch, and is the minimum-variance instance of the weighted batch mean. However, Bischak, et al. assume that the process $\{X_t\}$ is in steady state and that any initial transient observations have been deleted, and they use a different consistent estimator of ϕ. $\hat{\mu}$ can readily be computed for autoregressive processes of any order, and reduces to the GLS estimator derived by Snell and Schruben (1984) for the AR(1) process.

For autoregressive processes, we expect $\hat{\mu}$ to be nearly unbiased regardless of the initial observations. If the autocorrelation structure of the process is known, then $\hat{\mu}$ is exactly unbiased and has the same variance as the sample mean. If model (1) holds, and $\phi_1, \ldots, \phi_p$ are known and satisfy the stationarity condition of (4), then the MLE is exactly unbiased for the mean, $E(\hat{\mu}) = \mu$ $\forall \{X_1, \ldots, X_p\}$ and $Var(\hat{\mu}) = \sigma_\varepsilon \left[1 - \sum_{i=1}^{p} \phi_i \right]^{-2}$.

4 CONFIDENCE INTERVALS FOR STEADY STATE MEAN

In this section, we develop confidence intervals for the steady state for both single-replication and multiple-replication experiments. In order to develop a confidence interval procedure based on $\hat{\mu}$, we need to determine its distribution. We first show that the MLE is asymptotically normal, which suggests a single-replication interval

similar to Fishman (1978) and a multiple-replication procedure very much like that of the traditional replication-deletion method. Next we show that the MLE is asymptotically distributed as the index of a bivariate normal random variable, which suggests a more complex interval procedure which is likely more accurate in short series.

Single-replication intervals

LEMMA 1:

Let $\bar{X}_{n-p} = (n-p)^{-1}\sum_{t=p+1}^{n} X_t$ denote the p-truncated sample mean. Assume $\{X_t\}$ converges to a covariance stationary process satisfying (1). If $\hat{\mu}_n$ is the conditional least squares estimate (5) where p is fixed and finite, then $\hat{\mu}_n$ converges to the sample mean in probability

$$n^{1/2}(\hat{\mu}_n - \bar{X}_{n-p}) \to 0 \qquad (9)$$

Proof: We can rewrite () as the sample mean plus a bias correction:

$$\hat{\mu}_n = \bar{X}_{n-p} + \frac{\sum_{i=1}^{p}\hat{\phi}_i\left(\sum_{t=p-i+1}^{p} X_t - \sum_{t=n-i+1}^{n} X_t\right)}{(n-p)\left(1-\sum_{i=1}^{p}\hat{\phi}_i\right)} \qquad (10)$$

The bias correction is a finite sum of less than $2p^2$ terms of X_t. Since X_t is covariance stationary for large t with finite variance, it satisfies the stationarity condition (4), so the $1-\sum_{i=1}^{p}\hat{\phi}_i$ term in the denominator converges to a nonzero quantity in probability. Further, since the bias correction term is a finite sum of variables with finite mean and variance, and is divided by $n-p$, it converges to zero in probability as $n \to \infty$. Therefore, $n^{1/2}(\hat{\mu}_n - \bar{X}_{n-p}) \to 0$ in probability.

Corollary:

If (1) holds and satisfies the stationarity condition (4), and $\hat{\mu}_n$ is the maximum likelihood estimate of μ, then $\hat{\mu}$ is also asymptotically normal, $n^{1/2}(\hat{\mu}_n - \mu) \to N\left(0, \sigma_\varepsilon^2\left[1-\sum_{i=1}^{p}\phi_i\right]^{-2}\right)$. This follows from Equation (1) and from Yuan and Nelson, (1994), who note that $n^{1/2}(\bar{X}_n - \mu) \to N\left(0, \sigma_\varepsilon^2\left[1-\sum_{i=1}^{p}\phi_i\right]^{-2}\right)$.

Since the MLE has the same asymptotic variance as the sample mean and is based on an autoregressive model of the output process, a single-replication confidence interval similar to that of Fishman (1978) and Yuan and Nelson (1994):

$$\hat{\mu}_n \pm t_{1-\alpha/2,\hat{f}}\sqrt{\hat{Var}(\hat{\mu}_n)} \qquad (11)$$

where $t_{1-\alpha/2,\hat{f}}$ denotes the corresponding quantile of the student-t distribution, and $\hat{Var}(\hat{\mu}_n) = \hat{\sigma}_\varepsilon^2\left[1-\sum_{i=1}^{p}\hat{\phi}_i\right]^{-2}/(n-p)$. An estimate for the degrees of freedom is

$$\hat{f} = \frac{n(1-\sum\hat{\phi}_i)}{(1+2p)(1-\sum\hat{\phi}_i)+4\sum i\hat{\phi}_i}. \qquad (12)$$

The approximate estimator (12) is specifically derived for short series. However, in simulation experiments of AR(1) processes with $\phi_1 = 0.95$ and $n = 128$, this often yielded estimates of less than 1.0, leading to confidence intervals with tails heavier than even the Cauchy distribution. Here, we arbitrarily restrict $\hat{f} \geq 2$.

Since the MLE is the ratio of two random variables, $\hat{\phi}_0$ and $(1-\sum\hat{\phi}_i)$, and the variability of the denominator may be considerable at small sample sizes, a more accurate solution may be derived as follows. For convenience, let us write $\hat{\mu}_n = \hat{W}_{n,1}/\hat{W}_{n,2}$ where

$$\boldsymbol{W} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} \phi_0 \\ 1-\sum_{i=1}^{p}\phi_i \end{bmatrix}$$

and $\hat{\boldsymbol{W}}_n$ is defined similarly in terms of $\hat{\phi}_n$. From (7), we can see that $n^{1/2}(\hat{\boldsymbol{W}}_n - \boldsymbol{W}) \to N(0, \Sigma)$, since finite sums of asymptotically normal variables are themselves asymptotically normal. Therefore $\hat{\mu}_n$ is the index of an asymptotically bivariate normal random variable. Let $\boldsymbol{B} = \boldsymbol{A}^{-1}\sigma_\varepsilon^2$ with elements $\{b_{ij}\}$ be the limiting covariance matrix of $n(\hat{\phi}-\phi)$. We can write the limiting covariance matrix Σ as:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} b_{11} & -\sum_{i=2}^{p+1} b_{i1} \\ -\sum_{i=2}^{p+1} b_{i1} & \sum\sum_{i,j=2}^{p+1} b_{ij} \end{bmatrix} \qquad (13)$$

The following result is based on Fieller (1932), who derived the distribution of the ratio of two correlated non-central normal random variables. For a bivariate normal rv with mean $\sqrt{n}\boldsymbol{W}$ and covariance matrix Σ, the distribution of the index is asymptotically:

$$g_n(y) = \frac{k_1}{2(k_2 a)^{3/2}}\left(2\sqrt{k_2 a}\exp((-k_2 c)) + k_2 b\sqrt{\pi}\exp\left(-\frac{k_2}{4a}(4ac-b^2)\right)erf\left(\frac{k_2 b}{2\sqrt{k_2 a}}\right)\right) \qquad (14)$$

where

$$a = n\left[\left(\frac{y}{\sigma_1}\right)^2 - 2\rho\left(\frac{y}{\sigma_1}\right)\left(\frac{1}{\sigma_2}\right) + \left(\frac{1}{\sigma_2}\right)^2\right]$$

$$b = 2\sqrt{n}\left[\left(\frac{y}{\sigma_1}\cdot\frac{W_1}{\sigma_1}\right) - \rho\left(\frac{y}{\sigma_1}\cdot\frac{W_2}{\sigma_2} + \frac{W_1}{\sigma_1}\cdot\frac{1}{\sigma_2}\right) + \left(\frac{W_2}{\sigma_2^2}\right)\right]$$

$$c = \left[\left(\frac{W_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{W_1}{\sigma_1}\cdot\frac{W_2}{\sigma_2}\right) + \left(\frac{W_2}{\sigma_2}\right)^2\right]$$

$$k_1 = n\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-1}$$

$$k_2 = \frac{1}{2(1-\rho^2)}$$

and $G_n(y) = \int_{-\infty}^{y} g_n(t)\,dt$.

Let $f_n(y|\phi)$ denote the true finite-sample probability density function of $\hat{\mu}_n$. We can approximate $f_n(y|\phi)$ by $g_n(y|\phi)$. Similarly, we can approximate the cumulative distribution function, $F_n(y|\phi)$, by $G_n(y|\phi)$. For any fixed y, $g_n(y|\phi) \to f_n(y|\phi)$ and $G_n(y|\phi) \to F_n(y|\phi)$.

If $\{\hat{\phi}, \hat{\sigma}_\varepsilon^2\}$ are the maximum likelihood estimates, and $\hat{\mu}$ is defined, the finite sample distribution of $\hat{\mu}$ may be approximated as follows. An estimate of the covariance matrix, $\boldsymbol{B}$, is $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1}\hat{\sigma}^2$ where the elements of $\hat{\boldsymbol{A}}$ are

$$\hat{a}_{ij} = \begin{cases} \tilde{\gamma}_X(|i-j|) + \hat{\mu}^2; & i,j=2,3,\ldots,p+1 \\ 1; & i=j=1 \\ \hat{\mu}; & \text{otherwise} \end{cases} \quad . \qquad (15)$$

Estimates of the autocovariances, $\tilde{\gamma}(i)$, may be determined by solving the system of $p+1$ linear equations: $\tilde{\gamma}(i) - \hat{\phi}_1\tilde{\gamma}(|i-1|) - \ldots - \hat{\phi}_p\tilde{\gamma}(|i-p|) = 0$, $i = 1,\ldots,p$ and $\tilde{\gamma}(0) = \hat{\sigma}_\varepsilon^2$. The estimated covariance matrix $\hat{\boldsymbol{B}}$ may be substituted into (13) to obtain the estimated covariance matrix $\hat{\Sigma}$ and the estimated probability density and cumulative distribution functions, $\hat{g}_n(y)$ and $\hat{G}_n(y)$, respectively. This leads to the following $1-\alpha$ asymptotically valid confidence internal for μ:

$$\left[\hat{\mu}_n - \left(\hat{G}_n^{-1}\left(1-\frac{\alpha}{2}\right) - \hat{\mu}_n\right), \hat{\mu}_n + \left(\hat{\mu}_n - \hat{G}_n^{-1}\left(\frac{\alpha}{2}\right)\right)\right] \qquad (16)$$

Multiple-replication intervals

In multiple-replication experiments, we can let the data suggest the sampling distribution of $\hat{\mu}$. Consider R independently-seeded replications of the same simulation model running either in parallel on R processors, or in sequence on a single processor. Further assume that for each replication, the same number of observations are collected, avoiding the issue of ratio bias (Glynn and Heidelberger, 1992). For each replication, we can identify and estimate an autoregressive model and obtain an estimate for the mean, $\hat{\mu}(j)$ $j = 1,\ldots,R$. Since $\hat{\mu}$ is asymptotically normal, this suggests the following confidence interval for μ:

$$\bar{\hat{\mu}} \pm t_{1-\alpha/2,R-1}\sqrt{\hat{Var}(\bar{\hat{\mu}})} \qquad (17)$$

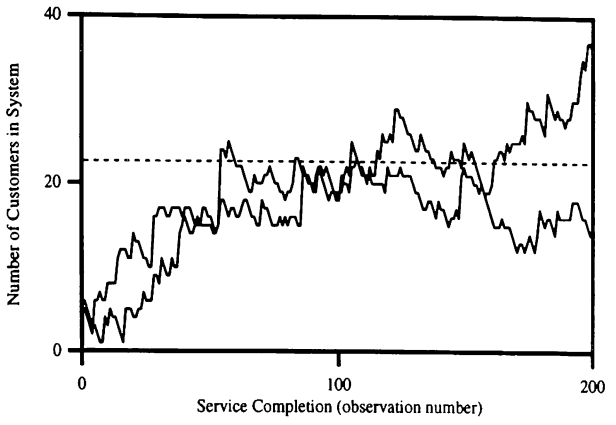
where $\bar{\hat{\mu}} = R^{-1}\sum_{j=1}^{R}\hat{\mu}(j)$ and $\hat{Var}(\bar{\hat{\mu}}) = R^{-1}(R-1)^{-1}\sum_{j=1}^{R}(\hat{\mu}(j) - \bar{\hat{\mu}})^2$. This is similar to the traditional confidence interval for the method of independent replications (see Law, 1991) except that it is based on the MLE $\hat{\mu}$ instead of the sample mean $\bar{X}$.

5 NUMERICAL EXPERIMENTS

The performance of the proposed estimator and confidence interval procedure is demonstrated empirically in a series of numerical experiments for the queuing network studied by Schruben (1982). The sequence of the number of entities in the system just before each customer arrival is shown in Figure 2 for two independent replications. On this graph is also superimposed the steady state mean, which is 22.40 ± 0.12 with 95% confidence, based on 100,000 observations. We see here that in short series, the initial transient period can be a substantial portion of the available data.

We compare the proposed procedure with the method of unweighted batch means (UBM) (see Law and Kelton, 1991) and weighted batch means (WBM). In single-replication experiments, MLE (F) indicates intervals based on Fieller's distribution (16) and MLE (N) indicates intervals based on the normal distribution (11); in multiple-replication experiments, intervals are computed by (17). We evaluate estimators on the basis of mean error (bias) and root-mean-square-error, and confidence intervals on the basis of coverage and mean interval width

FIGURE 1. Number of customers in system vs. time



In UBM, the series is optimally truncated to minimize the mean square error of the sample mean, as in Snell and Schruben (1984). For the series in Figure 1, the optimal truncation point is 36 when initialized empty and idle, and 0 when initialized in steady state. The entire observed series is always used for MLE and WBM since these are robust to initialization bias. For UBM and WBM, the batch size is 8 when $n = 128$ and 16 when $n \geq 256$. The autoregressive order is estimated by minimizing the finite-sample information criterion (FIC) (Broerson and Wensink,1993). The maximum candidate order p_{max}, is one unit less than the batch size since this is the largest value which can be chosen with WBM; p_{max} is therefore 7 when $n = 128$ and 15 when $n \geq 256$.

The output analysis procedures and process (1) simulation are written in C++ and compiled using the SparcWorks on a Sun workstation. Simulations are conducted under SIMAN IV on an IBM RS6000 workstation, and the results are exported to a text file for analysis.

Single replication Intervals:

In these experiments, series of length $n = \{128, 256, 512\}$ are generated for 1000 super-replications, and series of length $n = \{8192, 16384\}$ are generated for 200 super-replications. Table 1 summarizes performance of the estimators and interval procedures for single-replication experiments. First, in short simulations when the process is initialized empty and idle, MLE and WBM exhibit nearly the same bias, and are much less biased than even the optimally truncated sample mean in UBM, though MLE and WBM also exhibit considerably higher root-mean-square error than UBM. Second, in short samples initialized at steady state, all three estimators are practically unbiased, though MLE and WBM exhibit larger root-mean-square error for short series. Third, for large samples, all three estimators, UBM, WBM and MLE perform nearly identically in both mean error and root mean square error regardless of initial conditions. While optimal truncation can greatly minimize the root-mean-square error of the sample mean in short series, enough residual transient response remains that the truncated sample mean is still somewhat biased.

In addition to an estimate of the steady state mean, we also require an accurate assessment of how accurate the estimator is. Table 1 shows coverage frequencies and mean widths for nominally 90% confidence intervals. In short series initialized empty and idle, MLE(F) and MLE(N) yield coverages equal or very close to nominal coverages for all run lengths studied, even for very short run lengths $n = 128$. Further, MLE yields coverages which are much closer to nominal coverage than those obtained by UBM and WBM. In short series which are initialized from the steady state distribution, none of the three interval procedures attain nominal coverages, but MLE(F) and MLE(N) again yield coverages closer to

TABLE 1. Estimator Performance — single replication experiments

		Mean Error			Root Mean Square Error			Coverage Frequency				Mean Interval Width			
N	Initialized	UBM	WBM	MLE	UBM	WBM	MLE	UBM	WBM	MLE (F)	MLE (N)	UBM	WBM	MLE (F)	MLE(N)
128	Empty	-4.8	0.30	0.26	5.8	15.	15.	25%	71%	90%	91%	2.9	7.2	13.	12.
256	Empty	-2.1	-0.80	-0.80	3.1	2.4	2.4	40%	81%	93%	95%	2.0	4.3	7.1	7.2
512	Empty	-0.88	-0.48	-0.48	1.8	1.9	1.9	62%	81%	89%	89%	1.8	2,8	3.5	3.8
8192	Empty	-0.048	-0.036	-0.035	0.35	0.35	0.35	89%	89%	88%	88%	0.62	0.62	0.54	0.55
16392	Empty	-0.046	-0.039	-0.038	0.24	0.24	0.24	89%	89%	89%	89%	0.43	0.43	0.43	0.43
128	SSD	-0.32	0.28	0.29	2.7	4.0	3.9	56%	73%	77%	78%	2.0	3.9	5.3	5.1
256	SSD	-0.10	0.20	0.23	2.0	2.9	2.6	57%	78%	80%	81%	1.5	3.0	3.8	3.9
512	SSD	0.019	0.14	0.14	1.4	1.5	1.5	67%	81%	82%	84%	1.5	2.2	2.3	2.4
8192	SSD	-0.040	-0.025	-0.025	0.30	0.30	0.30	92%	91%	93%	93%	0.60	0.60	0.54	0.55
16392	SSD	-0.0057	-0.0029	-0.0027	0.24	0.24	0.24	91%	91%	90%	90%	0.43	0.43	0.43	0.43

nominal than those of UBM and WBM. In long series with $n \geq 8192$, all intervals yield coverages nearly equal to nominal. In every case, WBM also yields coverages closer to nominal than UBM, though MLE yields coverages closer to nominal than WBM.

The improved coverage of MLE(F) and MLE(N) come at the expense of wider intervals, just as the improved bias comes at the expense of increased root-mean-square error. While for very long series all the interval widths are nearly equal, for very short series, MLE(F) and MLE(N) yield intervals somewhat wider than WBM and much wider than UBM. However, this appears to be an unavoidable result of the short series. Interval-width is a valid measure of relative performance only among those methods which achieve nominal, or at least superior, coverage. Lastly, for this example, the simpler MLE(N) yields intervals nearly as accurate as MLE(F) on each of the four performance measures.

Multiple-Replication Intervals

Table 2 summarizes the performance of UBM, WBM, and MLE for multiple-replication confidence interval procedures. In these experiments, the total number of observations, N, is held constant, and are obtained by running R independent replications of length $n = N/R$. Intervals are computed as in (17), except for $R = 1$ where intervals are obtained using the corresponding single-replication procedures. Statistics are collected over 200 super-replications. Because of difficulty in generating and storing simulation results, these results are based on simulations of an AR model identified and estimated from an observed series of 100,000 observations of the model in Figure 1.

When all independent replications are initialized from the steady state distribution, all three estimators, UBM, WBM and MLE, result in confidence intervals with nominal coverage and comparable interval widths, as expected. However, initializing all replications from the steady state distribution is usually only possible by truncating excessive numbers of observations.

When each independent replication is initialized empty and idle, the MLE emerges as the superior estimator. As the available observations consist of more replications of shorter series, the optimally-truncated sample mean in UBM gradually becomes more biased, and the coverage of nominally 90% intervals based on the sample mean deteriorates rapidly. However, parallel intervals based on the MLE and WBM are much more robust to initialization bias. For $R = 32$, 90% intervals based on UBM yield coverages of 72%, whereas MLE and WBM yield about 86%. Further, the interval widths for all methods are nearly equal, regardless of whether the series are initialized empty or in steady state, or the number of replications into which available data is distributed. Even more extreme, for $R = 64$, nominally 90% intervals based on UBM yield coverages of 0%, whereas those based on MLE and WBM yield about 80%. Note again that for relatively few replications of long series, all methods yield nearly identical results, regardless of initial conditions.

These results indicate that when running multiple replications, there is little harm in using the MLE as the point estimator, and that in highly replicated series, there can be significant advantage. One would do almost as well using the WBM, but the MLE performs somewhat better for the example considered here.

6 SUMMARY

These numerical experiments demonstrate that the proposed estimator and distribution are robust to transient

TABLE 2. Estimator Performance — multiple replication experiments

		Mean Error			Root Mean Square Error			Coverage Frequency			Mean Interval Width		
N	Initialized	UBM	WBM	MLE	UBM	WBM	MLE	UBM	WBM	MLE	UBM	WBM	MLE
128	Empty	-.046	-.039	-.039	0.25	0.24	0.24	89%	89%	89%	0.43	0.43	0.39
256	Empty	-0.12	-0.025	-0.023	0.27	0.25	0.24	84%	88%	89%	0.41	0.41	0.41
512	Empty	-0.26	-0.074	-0.074	0.36	0.26	0.26	72%	87%	86%	0.41	0.41	0.41
8192	Empty	-1.33	-0.29	-0.27	1.37	0.87	0.76	0%	80%	81%	0.56	0.81	0.75
16392	Empty	-1.3	-0.36	-0.33	1.3	0.50	0.45	0%	64%	65%	0.39	0.52	0.48
128	SSD	-0.0056	-0.0030	-0.0027	0.24	0.24	0.24	90%	91%	90%	0.43	0.44	.039
256	SSD	-0.028	-0.016	-0.019	0.23	0.23	0.23	90%	90%	90%	0.40	0.40	0.40
512	SSD	-0.048	-0.032	-0.035	0.22	0.23	0.23	90%	90%	90%	0.39	0.41	0.41
8192	SSD	-0.032	-0.015	-0.015	0.24	0.24	0.24	89%	89%	89%	0.38	0.40	0.40
16392	SSD	-0.018	-0.018	-0.018	0.22	0.23	0.22	90%	90%	90%	0.43	0.43	0.39

behavior resulting from initial conditions, yielding estimates and confidence intervals which are comparable or superior to those of unweighted batch means and weighted batch means, even for extreme initializations. The confidence interval procedures presented here permit rapid evaluation of complex systems on the basis of steady state mean in short simulations. They can be usefully applied for providing decision support in evaluating many systems in real-time, where the cost of simulating long series of observations is prohibitively expensive. An areas for further investigation is rapid ranking and selection based on the MLE.

ACKNOWEDGMENTS

This research was partially supported by NASA under grant NAG-1-1469, and by the Department of Decision Sciences and Decision Systems, Rensselaer Polytechnic Institute.

REFERENCES

Bischak, D. P., W. D. Kelton, and S. M. Pollock. 1993. Weighted Batch Means For Confidence Intervals In Steady-State Simulations. Management Science, 39, 1002-1019.

Broerson, P. M. T., and H. E. Wensink. 1993. On Finite Sample Theory for Autoregressive Model Order Selection. IEEE Transactions on Signal Processing, 41, 194-204.

Fieller, E. C. 1932. The Distribution of the Index in a Normal Population. Biometrika, 24, 428-440.

Fishman, G. S. 1978 Principles of Discrete Event Simulation. John Wiley and Sons, New York.

Fuller, W. A. 1976. Introduction to Statistical Time Series. John Wiley and Sons, New York.

Glynn, P., and P. Heidelberger. 1992. Experiments with Initial Transient Deletion for Parallel, Replicated Steady State Simulations. Management Science, 38, 400-419.

Kelton, W. D., and A. M. Law. 1985. The Transient Behavior of the M/M/s Queue, with Implications for Steady-State Simulation. Operations Research, 33, 378-396.

Law, A. M. 1977. Confidence Intervals in Discrete Event Simulation: A Comparison of Replications and Batch Means. Naval Logistics Review Quarterly, 24, 667-678.

Law, A. M., and W. D. Kelton. 1991. Simulation Modeling and Analysis, 2nd ed. McGraw-Hill, New York.

Murray, J. R., and W. D. Kelton. 1988. The Transient Behavior of the M/E2/2 Queue and Steady-State Simulation. Computers & Operations Research, 15, 357-367.

Odoni, A. R., and E. Roth. 1983. An Empirical Investigation of the Transient Behavior of Stationary Queueing Systems. Operations Research, 31, 432-455.

Pawlikowski, K. 1990. Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions. ACM Computing Surveys, 22, 123-171.

Roth, E. 1994. The Relaxation Time Heuristic for the Initial Transient Problem in M/M/k Queueing Systems. European Journal of Operational Research, 72, 376-386.

Sargent, R. G., K. Kang, and D. Goldsman. 1992. An Investigation of Finite-Sample Behavior of Confidence Interval Estimators. Operations Research, 40, 898-913.

Schruben, L. 1982. Detecting Initialization Bias in Simulation Output. Operations Research, 30, 569-590.

Snell, M. and L. Schruben. 1984. Weighting Simulation Data to Reduce Initialization Effects. IIE Transactions, 17, 354-363.

Voss, P. A., J. Haddock, and T. R. Willemain. 1996. Estimating Steady State Mean in Short Simulations with Initial Transient. Technical Report 37-96-443. Dept. of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute. Troy, New York.

Yuan, M., and B. Nelson. 1994. Autoregressive-output-analysis methods revisited, Annals of Operations Research, 53, 391-418.

AUTHOR BIOGRAPHY

Pieter A. Voss is a Ph.D. Candidate in Operations Research and Statistics at Rensselaer. Research interests include simulation analysis and engineering pedagogy. As instructor of Engineering Economics, he received the Master Teaching Fellowship and Outstanding TA Award. He is president of the RPI Student Chapter of InfORMS. He has industrial experience with the Cedel Group, General Electric, the US Army, and Ingersoll-Rand.

Jorge Haddock is Professor of Industrial Engineering and Operations Research and Director of Undergraduate Programs in the department of Decision Sciences and Engineering Systems at Rensselaer. He received his Ph.D. in Industrial Engineering from Purdue University. His research interests are modeling of production and service systems, and analysis methodology. He received the Outstanding Young Industrial Engineer Award in 1990.

Thomas R. Willemain is Associate Professor of Decision Sciences and Engineering Systems at Rensselaer. He received the BSE from Princeton University and the SM and Ph.D. from Massachusetts Institute of Technology, all in Electrical Engineering. His research interests include time series analysis and forecasting.