
A Batch Means Methodology for Estimation of a Nonlinear Function of a Steady-State Mean

Author(s): David F. Munoz and Peter W. Glynn

Source: *Management Science*, Vol. 43, No. 8 (Aug., 1997), pp. 1121-1135

Published by: INFORMS

Stable URL: <http://www.jstor.org/stable/2634574>

Accessed: 21/07/2010 02:02

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=informs>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Management Science*.

A Batch Means Methodology for Estimation of a Nonlinear Function of a Steady-State Mean

David F. Muñoz • Peter W. Glynn

*Departamento de Administración, Instituto Tecnológico Autónomo de México,
Camino a Santa Teresa 930, México D.F. 10700*

Department of Operations Research, Stanford University, Palo Alto, California 94305

We study the estimation of steady-state performance measures from an \mathfrak{R}^d -valued stochastic process $Y = \{Y(t) : t \geq 0\}$ representing the output of a simulation. In many applications, we may be interested in the estimation of a steady-state performance measure that cannot be expressed as a steady-state mean r , e.g., the variance of the steady-state distribution, the ratio of steady-state means, and steady-state conditional expectations. These examples are particular cases of a more general problem—the estimation of a (nonlinear) function $f(r)$ of r . We propose a batch-means-based methodology that allows us to use jackknifing to reduce the bias of the point estimator. Asymptotically valid confidence intervals for $f(r)$ are obtained by combining three different point estimators (classical, batch means, and jackknife) with two different variability estimators (classical and jackknife). The performances of the point estimators are discussed by considering asymptotic expansions for their biases and mean squared errors. Our results show that, if the run length is large enough, the jackknife point estimator provides the smallest bias, with no significant increase in the mean squared error.

(Simulation Output Analysis; Steady-State Simulation; Nonlinear Estimation; Batch Means; Jackknifing)

1. Introduction

Performance measures of a stochastic system are usually expressed in terms of expected values or long-run averages. In most applications we consider systems that are stable in the sense that the \mathfrak{R}^d -valued stochastic process $Y = \{Y(s) : s \geq 0\}$, representing the output of our simulation, possesses a steady-state mean r , that is,

$$r(t) \stackrel{\text{def}}{=} \frac{1}{t} \int_0^t Y(s) ds \Rightarrow (r_1, r_2, \dots, r_d)^T = r, \quad (1)$$

where $r \in \mathfrak{R}^d$ is a constant and \Rightarrow denotes weak convergence (as $t \rightarrow \infty$ unless specified).

The vast majority of the existing literature on steady-state simulation focuses on estimation of the vector r , and construction of associated confidence regions. In this paper, however, our emphasis is on the development of general methods appropriate for handling estimation problems associated with performance measures that cannot be expressed as a steady-state mean (cf. Law and Kelton 1991: 285–287; Fishman and Moore 1979; Iglehart 1976; Shedler 1987).

EXAMPLE 1. Steady-state variance: Let us suppose that the stochastic process $X = \{X(t) : t \geq 0\}$ is an

\Re -valued stochastic process that has a steady-state distribution F , that is,

$$X(t) \Rightarrow X(\infty),$$

where $X(\infty)$ is a random variable with cumulative distribution function F . The steady-state variance of X is defined as

$$\text{Var}[X(\infty)] = E[X^2(\infty)] - (E[X(\infty)])^2 = r_2 - r_1^2, \quad (2)$$

where $r_2 = E[X^2(\infty)]$ and $r_1 = E[X(\infty)]$. Note that, if

$$\frac{1}{t} \left(\int_0^t X(s) ds, \int_0^t X^2(s) ds \right) \Rightarrow (r_1, r_2)^T,$$

then $r = (r_1, r_2)^T$ is a steady-state mean for the process $Y = \{Y(t) : t \geq 0\}$, where $Y(t) = (Y_1(t), Y_2(t))^T$, $Y_1(t) = X(t)$ and $Y_2(t) = X^2(t)$ for $t \geq 0$. Of course, it is obvious that $\text{Var}[X(\infty)]$ can be estimated as in (1), provided that one takes $Y(t) = (X(t) - r_1)^2$. Thus, in principle, estimation of the steady-state variance takes the form of the classical estimation problem (1). But, in reality, such an approach is infeasible, since the parameter $r_1 = E[X(\infty)]$ is unknown, and hence the process $Y = \{Y(t) : t \geq 0\}$ just constructed cannot be observed directly from the output of the simulation. Consequently, it is necessary to apply the "nonlinear" methods described in this paper.

EXAMPLE 2. Steady-state conditional expectations: Consider a stochastic process X that has a steady-state distribution F as in Example 1. One may be interested in a conditional expectation of the form

$$\begin{aligned} & E[g(X(\infty)) | X(\infty) \in A] \\ &= \frac{E[g(X(\infty))I(X(\infty) \in A)]}{E[I(X(\infty) \in A)]} = \frac{r_1}{r_2}, \end{aligned} \quad (3)$$

where g is a given real-valued function and I denotes the indicator function, that is, for any event B , $I(B)$ is 1 if event B occurs, and 0 otherwise. Note that, as in Example 1, $r = (r_1, r_2)^T$ can be viewed as the steady-state mean of a suitably defined stochastic process $Y = \{Y(t) : t \geq 0\}$.

EXAMPLE 3. Steady-state mean of a discrete-event stochastic system: Let the process X be the output process corresponding to a discrete-event simulation with state transition times $0 = \xi(0) < \xi(1) < \dots$ satisfying

$\xi(n) \rightarrow \infty$ a.s., as $n \rightarrow \infty$ a.s. (Glynn 1989), where "a.s." denotes convergence almost surely (see p. 29 of Chung 1974), and let g be a real-valued function. In a discrete-event simulation, the output process X has piecewise constant sample paths, so that it typically takes the form

$$X(t) = \sum_{n=0}^{\infty} X[\xi(n)]I[\xi(n) \leq t < \xi(n+1)].$$

Suppose that

$$\frac{1}{n} \sum_{i=0}^{n-1} g\{Y[\xi(i)]\}[\xi(i+1) - \xi(i)] \rightarrow r_1 \quad \text{a.s.}, \quad (4)$$

and

$$\frac{1}{n} \xi(n) \rightarrow r_2 \quad \text{a.s.}, \quad (5)$$

as $n \rightarrow \infty$, where r_1, r_2 are finite constants, and $r_2 > 0$. The above laws of large numbers hold in great generality in the discrete-event systems setting. In any case, under these assumptions, it follows that

$$\frac{1}{t} \int_0^t g[X(s)] ds \rightarrow \frac{r_1}{r_2} \quad \text{a.s.}$$

The proof of this result is given in Glynn and Iglehart (1988). Therefore, the estimation of the steady-state mean of a discrete-event simulation can be viewed as the estimation of the ratio

$$\alpha = \frac{r_1}{r_2}.$$

The above ratio representation of a steady-state mean is particularly advantageous in computational settings in which the simulation is implemented asynchronously, so that the "natural" time scale for collecting observations is that based on state transition epochs.

The estimation of the performance measures of Examples 1, 2, and 3 is a particular case of a more general problem. The general problem we address in this paper is the estimation of a (possibly nonlinear) function $f(r)$ of a (multivariate) steady-state mean r . The classical point estimator for the performance measure $f(r)$ is the evaluation $f[r(t)]$ of the function f at the sample average of the process $r(t)$ (where t is the run length). In a system that possesses a steady-state mean, the sample

average $r(t)$ is a consistent estimator for the steady-state mean r , so that $f[r(t)]$ is a consistent estimator for $f(r)$ (under the assumption of continuity of f). The main difficulty in the estimation of $f(r)$ is that, even for independent and identically distributed (i.i.d.) observations, in general we are not able to produce unbiased estimators (if f is nonlinear, the classical estimator $f[r(t)]$ is biased (Miller (1974))). For this reason, different estimation procedures will be compared in this paper according to the magnitude of the bias that they introduce and/or their mean squared errors.

In addition, the construction of confidence intervals for such nonlinear parameters requires appropriate methodology. While classical statistical methods apply to the estimation of $f(r)$ when r can be expressed as the mean of a vector-valued terminating simulation (Serfling 1980), this problem has not heretofore been addressed in the steady-state context. One of the major contributions of this paper is the development of a mathematically rigorous and asymptotically valid confidence interval procedure to produce confidence intervals for $f(r)$. Specifically, we propose a batch-means-based methodology that produces asymptotically valid confidence intervals under the assumption that a functional central limit theorem (FCLT) holds (see §2). The main advantages of our proposed methodology are its robustness (see, for example, Glynn and Iglehart 1990 and §1.2.3 of Muñoz 1991), and the simplicity of its implementation.

We start in §2 by describing the mathematical framework that underlies the study of our proposed methodology. In particular, we state our FCLT assumption concerning the stochastic process representing the output of the simulation. In §3, we describe our proposed methodology. Assessing the variability plays a key role in the construction of confidence intervals. We are interested not only in the point estimation of $f(r)$, but also in the assessment of the variability of the point estimator. We exploit the batch means method, which has demonstrated good performance in the estimation of the steady-state parameter r when f is linear. Based on the batch means method, we consider three consistent estimators for $f(r)$ (including a jackknife point estimator (cf. Miller 1964)) and two variability estimators (including a jackknife variability estimator). Under an FCLT

assumption, we obtain asymptotically valid confidence intervals for $f(r)$ based on these estimators. In §4, we discuss the performances of the point estimators given in §3 by considering asymptotic expansions for their biases and mean squared errors. In §5, we present and discuss experimental results from the application of our methodology to estimate a nonlinear function of a steady-state mean.

2. Mathematical Framework

As discussed in the Introduction, it is necessary that the output process Y satisfy the law of large numbers (1), in order that the steady-state estimation problem be well defined. However, the development of a confidence interval methodology requires making additional assumptions that permit one to describe the variability of the estimator $r(t)$ about the steady-state mean r . In particular, a standard assumption that (implicitly) underlies much of the existing steady-state simulation methodology is that $r(t)$ satisfies a multivariate form of the central limit theorem (CLT), namely, that there exists a $d \times d$ nonsingular matrix G such that

$$t^{1/2}(r(t) - r) \Rightarrow GN_d(0, I), \tag{6}$$

where $N_d(0, I)$ denotes the normal d -variate distribution with mean 0 and covariance matrix I (the identity). Note that $\Theta = GG^T$ is the covariance matrix of the limiting normal random vector appearing in the right-hand side of (6).

It turns out that the methodology that we shall propose here requires a slightly stronger type of assumption. Set

$$\bar{Y}_u(t) = \frac{1}{u} \int_0^{ut} Y(s) ds,$$

and

$$X_u(t) = u^{1/2}(\bar{Y}_u(t) - rt), \quad 0 \leq t \leq 1.$$

Recall that a d -dimensional standard Brownian motion is a stochastic process $B = \{B(t) : t \geq 0\}$ with stationary independent increments such that $B(t)$ has a $N_d(0, tI)$ distribution. We shall demand the following so-called functional central limit theorem (FCLT) for Y to be valid:

ASSUMPTION 1. *There exists a nonsingular $d \times d$ matrix G such that*

$$X_u \Rightarrow GB,$$

as $u \rightarrow \infty$ (in the topology of weak convergence in $C^d[0, 1]$, the space of \mathbb{R}^d -valued continuous functions defined on $[0, 1]$; see Ethier and Kurtz 1986 for additional discussion).

The reason that the ordinary CLT (6) typically holds for a steady-state simulation depends on the fact that observations taken from Y that are widely separated in time are approximately independent and identically distributed. As a consequence, $r(t)$ behaves very much like an average of i.i.d. random vectors, and one can therefore expect a CLT to hold. The same independence argument leads naturally to the additional structure associated with the FCLT required in Assumption 1. From a mathematical viewpoint, FCLT theorems have been established for Markov processes in discrete and continuous time and stationary processes satisfying so-called "mixing conditions"; see Glynn and Iglehart (1990) and Muñoz (1991) for additional details.

We note that Assumption 1 implies the law of large numbers (1). Thus, the steady-state estimation problem is always well defined under Assumption 1. In addition, we remark that while the mathematical discussion of this paper will focus exclusively on continuous-time output processes Y , any discrete-time output process $Z = \{Z_n : n \geq 0\}$ can be incorporated into our framework by setting $Y(t) = Z_{\lfloor t \rfloor}$, where $\lfloor t \rfloor$ is the integer part of t .

ASSUMPTION 2. *The function f is differentiable in a neighborhood of r .*

As we can see, Assumptions 1 and 2 together guarantee that $f[r(t)]$ is a consistent estimator for $f(r)$.

3. Methodology

3.1. Point Estimators

We have already seen that $f[r(t)]$ is a consistent (point) estimator for $f(r)$. Bias in $f[r(t)]$ is a consequence of the presence of initial transient effects, as well as the nonlinearities inherent in f . Many algorithms have been proposed in the literature for determining an appropriate "deletion time," before which all simulation output

is discarded; see for example, Wilson and Pritsker (1978). By using such methods, one can expect that the resulting point estimator $r(t)$ is then approximately unbiased as an estimator of r . Assuming that such an initial bias deletion algorithm has been implemented, the only significant bias effects on $f[r(t)]$ must then ensue from the nonlinearity of f . The estimation techniques that we shall consider in this paper are designed exclusively to deal with this "nonlinearity bias."

As mentioned in the Introduction, we will be using a batch-means-type methodology to produce confidence intervals for $f(r)$. Computational considerations therefore suggest that it is especially convenient if our bias adjustments take advantage of the specific batch structure used. In view of this, suppose that we now subdivide the run length into m batches of equal length. Set

$$\bar{X}_i(t) = \frac{m}{t} \int_{(i-1)t/m}^{it/m} Y(s) ds, \quad i = 1, 2, \dots, m. \quad (7)$$

We shall consider three different point estimators for $f(r)$:

(i) Classical estimator:

$$f[r(t)] = f \left[\frac{1}{t} \int_0^t Y(s) ds \right]. \quad (8)$$

(ii) Batch means estimator:

$$\bar{f}_m(t) = \frac{1}{m} \sum_{i=1}^m f[\bar{X}_i(t)]. \quad (9)$$

(iii) Jackknife estimator:

$$\alpha_m^J(t) = \frac{1}{m} \sum_{i=1}^m J_i^m(t) \quad (10)$$

where

$$J_i^m(t) = mf[r(t)] - (m - 1)f[\bar{X}_i^m(t)]$$

and

$$\bar{X}_i^m(t) = \frac{1}{m - 1} \sum_{j \neq i} \bar{X}_j(t).$$

Note that all three estimators coincide in the case where f is a linear function (i.e., when we are estimating a steady-state mean). The jackknife estimator was first introduced by Quenouille (1956) as a method of last resort when an unbiased estimator for a parameter is

not available. As was shown by Miller (1964, 1974), if X_1, X_2, \dots, X_m are i.i.d. observations, the bias of $f(\bar{X}_m)$, where \bar{X}_m is the sample mean of the X_i 's, is of the order of m^{-1} , whereas the bias of the jackknife estimator is of the order of m^{-2} . As we are going to see in §4, under Assumption 1, the jackknife estimator (as defined in (10)) has a bias of lower order than that of the classical estimator (as defined in (8)).

To obtain an asymptotic confidence interval for $f(r)$, we need to derive a CLT for the point estimator we are considering. It turns out that Assumption 1 allows us to obtain a CLT for each of the point estimators proposed. To establish our CLTs, we first consider a joint CLT for the batch means.

PROPOSITION 1. *Let $m \geq 1$ be fixed. Then under Assumption 1 we have*

$$\left(\frac{t}{m}\right)^{1/2} \begin{pmatrix} \bar{X}_1(t) - r \\ \bar{X}_2(t) - r \\ \vdots \\ \bar{X}_m(t) - r \end{pmatrix} \Rightarrow G_m N_{dm}(0, I),$$

where

$$G_m = \begin{pmatrix} G & 0 & \dots & 0 \\ 0 & G & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G \end{pmatrix}. \tag{11}$$

PROOF. We can view each batch of the simulation as an increment of the averaged cumulative process, by considering the function $\Lambda_i^m : C^d[0, 1] \rightarrow C^d[0, 1]$ defined by

$$(\Lambda_i^m x)(t) = x[(i + t)/m] - x(i/m), \\ 0 \leq i \leq m - 1, \quad 0 \leq t \leq 1.$$

Then, if we define $\Lambda^m : C^d[0, 1] \rightarrow R^{dm}$ such that

$$\Lambda^m(x) = \begin{pmatrix} (\Lambda_0^m x)(1) \\ (\Lambda_1^m x)(1) \\ \vdots \\ (\Lambda_{m-1}^m x)(1) \end{pmatrix}, \quad x \in C^d[0, 1],$$

we have that $\Lambda^m(GB)$ is distributed as $G_m N_{dm}(0, I)/m^{1/2}$.

On the other hand,

$$\begin{aligned} (\Lambda_i^m X_u)(1) &= u^{1/2} \left(\bar{Y}_u[(i + 1)/m] - \bar{Y}_u(i/m) - \frac{r}{m} \right) \\ &= \frac{u^{1/2}}{m} \left(m \int_{i/m}^{(i+1)/m} Y(s) ds - r \right) \\ &= \frac{u^{1/2}}{m} [\bar{X}_{i+1}(u) - r], \end{aligned}$$

so that

$$\Lambda^m(X_u) = \frac{u^{1/2}}{m} \begin{pmatrix} \bar{X}_1(u) - r \\ \bar{X}_2(u) - r \\ \vdots \\ \bar{X}_m(u) - r \end{pmatrix}. \tag{12}$$

Since Λ^m is a continuous mapping, the conclusion follows from (12), Assumption 1, and the continuous mapping theorem (Corollary 1.9 of Ethier and Kurtz 1986). □

Recall Assumption 2 concerning the differentiability of f . In order to derive our CLTs, we consider the first-order Taylor series expansion:

$$\begin{aligned} f(x) &= f(r) + [\nabla f(r)]^T(x - r) \\ &\quad + k(x - r), \quad x \in \mathfrak{R}^d, \end{aligned} \tag{13}$$

where the function $k : \mathfrak{R}^d \rightarrow \mathfrak{R}$ satisfies

$$\lim_{u \rightarrow 0} \frac{k(u)}{\|u\|} = 0, \tag{14}$$

and $\|u\|$ denotes the Euclidean norm of $u \in \mathfrak{R}^n$. From (13), an expansion for $f[r(t)]$ is given by

$$f[r(t)] = f(r) + [\nabla f(r)]^T(r(t) - r) + k(r(t) - r). \tag{15}$$

Similarly, from (9) and (13), we have

$$\begin{aligned} \bar{f}_m(t) &= f(r) + [\nabla f(r)]^T(r(t) - r) \\ &\quad + \frac{1}{m} \sum_{i=1}^m k(\bar{X}_i(t) - r). \end{aligned} \tag{16}$$

Further, (9), (10), (13), and some algebra yield

$$\begin{aligned} \alpha_m^l(t) &= f(r) + [\nabla f(r)]^T(r(t) - r) + mk(r(t) - r) \\ &\quad - \frac{m-1}{m} \sum_{i=1}^m k(\bar{X}_i^m(t) - r). \end{aligned} \tag{17}$$

As we see from (15), (16), and (17), the linear term in our expansions is the same for all three point

estimators. A CLT for the linear term can be derived from the CLT (6).

LEMMA 1. Under the CLT (6), we have

$$t^{1/2}[\nabla f(r)]^T(r(t) - r) \Rightarrow \sigma N(0, 1),$$

where $\sigma^2 = [\nabla f(r)]^T G G^T \nabla f(r)$.

PROOF. Since the function $w : \mathfrak{R}^d \rightarrow \mathfrak{R}$ defined by $w(x) = [\nabla f(r)]^T x$, $x \in \mathfrak{R}^d$, is continuous, the conclusion follows from the continuous mapping theorem. \square

Now, if we scale (15), (16), and (17) by $t^{1/2}$, we can show that the remainder terms converge weakly to 0. Consider, for example, (16). If we let

$$k_1(u) = \begin{cases} k(u)/\|u\|, & u \neq 0 \\ 0, & u = 0, \end{cases}$$

since $k(0) = 0$, we have

$$t^{1/2}k(\bar{X}_i(t) - r) = t^{1/2}\|\bar{X}_i(t) - r\|k_1(\bar{X}_i(t) - r), \quad i = 1, 2, \dots, m. \quad (18)$$

Under Assumption 1, it follows from Proposition 1 that $\|\bar{X}_i(t) - r\| \Rightarrow 0$. Note also that (14) implies that $k_1(u)$ is continuous at $u = 0$. Hence, if f is differentiable at r , from (14) and the continuous mapping theorem we have that

$$k_1(\bar{X}_i(t) - r) \Rightarrow 0, \quad i = 1, 2, \dots, m. \quad (19)$$

Now, from Proposition 1 and the continuous mapping theorem we have

$$t^{1/2}\|\bar{X}_i(t) - r\| \Rightarrow m^{1/2}\|GN_d(0, I)\|.$$

Then, from (18) and the converging together principle (cf. Billingsley 1986) we have

$$t^{1/2}k(\bar{X}_i(t) - r) \Rightarrow 0, \quad i = 1, 2, \dots, m,$$

so that from the converging together principle we have

$$\frac{t^{1/2}}{m} \sum_{i=1}^m k(\bar{X}_i(t) - r) \Rightarrow 0. \quad (20)$$

Similarly, we can show that, under Assumption 1 we have

$$t^{1/2}k(r(t) - r) \Rightarrow 0, \quad (21)$$

and

$$\frac{t^{1/2}(m-1)}{m} \sum_{i=1}^m k(\bar{X}_i^m(t) - r) \Rightarrow 0. \quad (22)$$

From (20)–(22), (15)–(17), and Lemma 1 we have the following proposition.

PROPOSITION 2. Under Assumptions 1 and 2, we have

$$t^{1/2}[\hat{f}_m(t) - f(r)] \Rightarrow \sigma N(0, 1),$$

where $\hat{f}_m(t)$ can be $f[r(t)]$, $\bar{f}_m(t)$ or $\alpha_m^J(t)$.

From the last proposition we see that if σ is consistently estimated (e.g., using the regenerative method), we can obtain asymptotic confidence intervals for $f(r)$. To be more precise, if $s(t)$ is a consistent estimator for σ and $\sigma > 0$, an asymptotic $100(1 - \delta)\%$ confidence interval for $f(r)$ is given by

$$\left[\hat{f}_m(t) - z_\delta \frac{s(t)}{t^{1/2}}, \hat{f}_m(t) + z_\delta \frac{s(t)}{t^{1/2}} \right],$$

where z_δ is the constant chosen so that $P[N(0, 1) \leq z_\delta] = 1 - \delta/2$.

We concentrate our attention on the case where σ is not consistently estimated. That is, we will try to cancel out the constant σ by scaling the process appropriately. For the batch means method, the proper scaling depends on the sample variance of the batch means. By following a similar approach, we scale the process by considering variability estimators that are sample variances obtained from the batches. We define our variability estimators in §3.2.

3.2. Variability Estimators

In this subsection, we assess the variability of the point estimators $\hat{f}_m(t)$. The following variability estimators can be proposed:

(i) The batch means variability estimator:

$$S_m^2(t) = \frac{1}{m-1} \sum_{i=1}^m \{f[\bar{X}_i(t)] - \bar{f}_m(t)\}^2. \quad (23)$$

(ii) The jackknife variability estimator:

$$S_J^2(t) = \frac{1}{m-1} \sum_{i=1}^m [J_i^m(t) - \alpha_m^J(t)]^2. \quad (24)$$

The batch means variability estimator is the sample variance of the $f[\bar{X}_i(t)]$'s and the jackknife variability estimator is the sample variance of the pseudovalues

$J_i^m(t)$. One important property of the jackknife variability estimator that holds in the i.i.d. case is known as the Efron-Stein inequality (cf. Efron and Stein 1981; Karlin and Rinott 1982), which states that the jackknife variability estimator overestimates the variance of a nonlinear function of the sample mean. This result suggests that confidence intervals based on the jackknife variability estimator tend to provide larger expected half-width than those based on the batch means variability estimator.

As we are going to see in the next set of propositions, under Assumption 1 we can obtain CLTs for the point estimators proposed in §3.1 by considering either of these variability estimators.

THEOREM 1. *Under Assumptions 1 and 2, we have*

$$\frac{\hat{f}_m(t) - f(r)}{S_m(t)/\sqrt{m}} \Rightarrow t_{(m-1)},$$

where $\hat{f}_m(t)$ can be $f[r(t)]$, $\bar{f}_m(t)$ or $\alpha_m^I(t)$, $S_m(t)$ is defined in (23) and $t_{(m-1)}$ denotes the Student- t distribution with $m - 1$ degrees of freedom.

PROOF. We first consider the case $\hat{f}_m(t) = \bar{f}_m(t)$. Let us denote

$$f_m(t) = \begin{pmatrix} f[\bar{X}_1(t)] - f(r) \\ f[\bar{X}_2(t)] - f(r) \\ \vdots \\ f[\bar{X}_m(t)] - f(r) \end{pmatrix}.$$

It follows from (13), (20), Proposition 1, and the converging together principle that

$$\left(\frac{t}{m}\right)^{1/2} f_m(t) \Rightarrow \sigma N_m(0, I). \tag{25}$$

Let $A = \{x \in \mathfrak{R}^m : x_1 = x_2 = \dots = x_m\}$ and consider the function $w : \mathfrak{R}^m \rightarrow \mathfrak{R}$ such that

$$w(x) = \begin{cases} \frac{\bar{x}_m}{s(x)/\sqrt{m}}, & x \in \mathfrak{R}^m - A, \\ 0, & x \in A, \end{cases}$$

where, for $x = (x_1, x_2, \dots, x_m)^T \in \mathfrak{R}^m$,

$$\bar{x}_m = \frac{\sum_{i=1}^m x_i}{m} \quad \text{and}$$

$$s^2(x) = \frac{\sum_{i=1}^m (x_i - \bar{x}_m)^2}{m - 1}.$$

Then w is a continuous function on $\mathfrak{R}^m - A$. Also, since $\sigma > 0$, $w(\sigma N_m(0, I))$ follows a $t_{(m-1)}$ distribution, and $P[\sigma N_m(0, I) \in A] = 0$. Therefore, it follows from the continuous mapping theorem that

$$w[(t/m)^{1/2} f_m(t)] \Rightarrow t_{(m-1)}. \tag{26}$$

Now we consider the cases $\hat{f}_m(t) = f[r(t)]$ or $\alpha_m^I(t)$. Let us define the function $u : \mathfrak{R}^m \rightarrow \mathfrak{R}$ by

$$u(x) = \frac{(m - 1)s^2(x)}{\sigma^2}.$$

Then, $u[\sigma N_m(0, I)]$ follows a $\chi_{(m-1)}^2$ distribution, so that, from (25) and the continuous mapping theorem we obtain

$$u[(t/m)^{1/2} f_m(t)] = \frac{t(m - 1)S_m^2(t)}{m\sigma^2} \Rightarrow \chi_{(m-1)}^2. \tag{27}$$

Therefore, since $P[\chi_{(m-1)}^2 > 0] = 1$, we have

$$\left(\frac{m}{t}\right)^{1/2} S_m^{-1}(t) \Rightarrow \frac{(m - 1)^{1/2}}{\sigma} \chi_{(m-1)}^{-1}, \tag{28}$$

where

$$\frac{(m - 1)^{1/2}}{\sigma} \chi_{(m-1)}^{-1}$$

is a proper random variable (since $\sigma > 0$). Now, from (28), (20)–(22), and the converging together principle, we have

$$\frac{\sum_{i=1}^m k(\bar{X}_i(t) - r)}{mS_m(t)/\sqrt{m}} \Rightarrow 0, \tag{29}$$

$$\frac{k(r(t) - r)}{S_m(t)/\sqrt{m}} \Rightarrow 0, \quad \text{and} \tag{30}$$

$$\frac{(m - 1) \sum_{i=1}^m k(\bar{X}_i^m(t) - r)}{mS_m(t)/\sqrt{m}} \Rightarrow 0. \tag{31}$$

Hence, the conclusion follows from (29)–(31), (26), and the converging together principle. \square

THEOREM 2. *Under Assumptions 1 and 2, we have*

$$\frac{\hat{f}_m(t) - f(r)}{S_f(t)/\sqrt{m}} \Rightarrow t_{(m-1)}, \quad t[S_f^2(t) - S_m^2(t)] \Rightarrow 0. \quad (33)$$

where $S_f(t)$ is defined in (24), and $\hat{f}_m(t)$ can be $f[r(t)]$, $\bar{f}_m(t)$ or $\alpha_m^l(t)$.

PROOF. From (13) and (10) we have

$$\begin{aligned} J_i^m(t) - \alpha_m^l(t) &= [\nabla f(r)]^T [\bar{X}_i(t) - r(t)] \\ &+ \frac{m-1}{m} \sum_{j=1}^m [\nabla f(r)]^T [\bar{X}_j^m(t) - r] \\ &- (m-1)k(\bar{X}_i^m(t) - r), \end{aligned}$$

and

$$\begin{aligned} f[\bar{X}_i(t)] - \bar{f}_m(t) &= [\nabla f(r)]^T [\bar{X}_i(t) - r(t)] \\ &+ k(\bar{X}_i(t) - r) - \frac{1}{m} \sum_{j=1}^m k(\bar{X}_j(t) - r), \end{aligned}$$

$i = 1, 2, \dots, m$, so that

$$\begin{aligned} J_i^m(t) - \alpha_m^l(t) &= f[\bar{X}_i(t)] - \bar{f}_m(t) \\ &+ \epsilon_i(t), \quad i = 1, 2, \dots, m, \end{aligned}$$

where

$$\begin{aligned} \epsilon_i(t) &= -(m-1) \left(k(\bar{X}_i^m(t) - r) - \frac{1}{m} \sum_{j=1}^m k(\bar{X}_j^m(t) - r) \right) \\ &- k(\bar{X}_i(t) - r) + \frac{1}{m} \sum_{j=1}^m k(\bar{X}_j(t) - r), \\ &i = 1, 2, \dots, m, \end{aligned}$$

so that

$$\begin{aligned} S_f^2(t) &= S_m^2(t) + \frac{2}{m-1} \sum_{i=1}^m \epsilon_i(t) \{f[\bar{X}_i(t)] \\ &- \bar{f}_m(t)\} + \frac{1}{m-1} \sum_{i=1}^m \epsilon_i^2(t). \quad (32) \end{aligned}$$

Under Assumption 1, we see from (20) and (22) that $t^{1/2}\epsilon_i(t) \Rightarrow 0, i = 1, 2, \dots, m$. Also, from (25) we see that $t^{1/2}(f[\bar{X}_i(t)] - \bar{f}_m(t))$ converges weakly to a finite random variable as $t \rightarrow \infty, i = 1, 2, \dots, m$. Therefore,

$$t \left[2 \sum_{i=1}^m \epsilon_i(t) \{f[\bar{X}_i(t)] - \bar{f}_m(t)\} + \sum_{i=1}^m \epsilon_i^2(t) \right] \Rightarrow 0.$$

Hence, from (32) we obtain

Now, since

$$P[\chi_{(m-1)}^2 > 0] = 1,$$

we have from (27) that

$$[tS_m^2(t)]^{-1} \Rightarrow \left(\frac{m\sigma^2\chi_{(m-1)}^2}{m-1} \right)^{-1}.$$

Therefore, from (33) and the converging together principle we have

$$\frac{S_f^2(t)}{S_m^2(t)} \Rightarrow 1,$$

so that

$$\frac{S_m(t)}{S_f(t)} \Rightarrow 1. \quad (34)$$

The conclusion follows from (34), Theorem 1, and the converging together principle. \square

Theorems 1 and 2 form the core of our main results on confidence interval construction. Together, they state that any of our three point estimators, in conjunction with either of our two variability estimators, produce asymptotically valid confidence intervals for $f(r)$. To be more precise, if $\sigma^2 > 0$, an asymptotic $100(1 - \delta)\%$ confidence interval for $f(r)$ is given by:

$$\left[\hat{f}_m(t) - t_{(\delta, m-1)} \frac{\hat{S}_m(t)}{m^{1/2}}, \hat{f}_m(t) + t_{(\delta, m-1)} \frac{\hat{S}_m(t)}{m^{1/2}} \right], \quad (35)$$

where $t_{(\delta, m-1)}$ is the constant chosen so that $P[t_{(m-1)} \leq t_{(\delta, m-1)}] = 1 - \delta/2$ and $\hat{S}_m(t)$ can be $S_m(t)$ or $S_f(t)$.

4. Asymptotic Bias and Mean Squared Error Expansions

Let us suppose that $\hat{\theta}$ is a point estimator for a parameter θ . There are two important measures that allow us to judge the accuracy and the precision, respectively, of $\hat{\theta}$ in estimating θ . These measures are the bias of $\hat{\theta}$:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta} - \theta], \quad (36)$$

and the mean squared error of $\hat{\theta}$:

$$\text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + \text{Bias}^2(\hat{\theta}). \quad (37)$$

In this section, we compare the long-run performance

of the three point estimators defined in §3.1, according to their asymptotic biases and their asymptotic mean squared errors. Because of the fact that all three candidate estimators are reasonable from a confidence interval viewpoint, the goal here is to study these “second-order” issues with the intent of using this second-order structure to differentiate among the three candidates. We start by obtaining asymptotic bias expansions.

4.1. Asymptotic Bias Expansions

As described in §3.1, the expansions for $\bar{f}_m(t)$, $f[r(t)]$, and $\alpha_m^J(t)$ based on first-order Taylor series expansions differ only in their residuals, so that the expansions do not allow us to compare the magnitude of the bias corresponding to each estimator. To analyze the bias, we need to consider a second-order Taylor series expansion. Let us suppose that f is twice differentiable at r . A second-order Taylor series expansion for $f(x)$ is given by

$$f(x) = f(r) + [\nabla f(r)]^T(x - r) + \frac{1}{2}(x - r)^T H(x - r) + w(x - r), \quad (38)$$

where $w : \mathfrak{R}^d \rightarrow \mathfrak{R}$ satisfies

$$\lim_{u \rightarrow 0} \frac{w(u)}{\|u\|^2} = 0,$$

and $H = (h_{ij})$ is the Hessian matrix of f evaluated at $x = r$, that is,

$$h_{ij} = \left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=r}, \quad i, j = 1, 2, \dots, d.$$

Then an expansion for $f[r(t)]$ is given by

$$f[r(t)] = f(r) + [\nabla f(r)]^T(r(t) - r) + \frac{1}{2}(r(t) - r)^T H(r(t) - r) + w(r(t) - r). \quad (39)$$

Similarly, we obtain

$$\begin{aligned} \bar{f}_m(t) &= f(r) + [\nabla f(r)]^T(r(t) - r) \\ &+ \frac{1}{2m} \sum_{i=1}^m (\bar{X}_i(t) - r)^T H(\bar{X}_i(t) - r) \\ &+ \frac{1}{m} \sum_{i=1}^m w(\bar{X}_i(t) - r), \end{aligned} \quad (40)$$

$$\begin{aligned} \alpha_m^J(t) &= f(r) + [\nabla f(r)]^T(r(t) - r) \\ &+ \frac{m}{2} (r(t) - r)^T H(r(t) - r) \\ &- \frac{m-1}{2m} \sum_{i=1}^m (\bar{X}_i^m(t) - r)^T H(\bar{X}_i^m(t) - r) \\ &+ mw(r(t) - r) - \frac{m-1}{m} \sum_{i=1}^m w(\bar{X}_i^m(t) - r). \end{aligned} \quad (41)$$

To compare the magnitude of the bias that each estimator introduces, we note that, under Assumption 1, the convergence rate of the residual terms in (39)–(41) is of order $n^{-3/2}$, whereas it is of order n^{-1} for the quadratic term. Therefore, the bias from the second-order term will dominate the bias from the residuals for large n , so that we can base our analysis on the quadratic term. To precisely describe our results, however, we require some control on the growth of f . To that end, we adopt the approach of Glynn and Heidelberger (1989) by imposing a regularity condition on the growth of f .

DEFINITION 1. Let $f : \mathfrak{R}^d \rightarrow \mathfrak{R}$. We say that f is polynomially dominated to degree q ($q \geq 0$) if there exist constants A and D such that

$$|f(x)| \leq A + D\|x\|^q, \quad x \in \mathfrak{R}^d.$$

Note that if f is bounded, then f is polynomially dominated to degree 0. Also, if all the partial derivatives of f of order q are globally bounded, then f is polynomially dominated to degree q . The next theorem provides asymptotic expansions for the bias of the point estimators proposed in §3.1.

THEOREM 3. Suppose that the process Y satisfies Assumption 1, and f is polynomially dominated to degree q ($q \geq 0$) and differentiable at r . Let the number of batches $m \geq 1$ be fixed, and $p = \max\{2, q\}$. If Y is a strictly stationary stochastic process, and there exists $t_0 > 0$ such that $\{t^{p/2}\|r(t) - r\|^p : t \geq t_0\}$ is uniformly integrable, then

$$\text{Bias}\{f[r(t)]\} = \frac{b}{t} + o(t^{-1}), \quad (42)$$

as $t \rightarrow \infty$,

$$\text{Bias}[\bar{f}_m(t)] = \frac{mb}{t} + o(t^{-1}), \tag{43}$$

and

$$\text{Bias}[\alpha'_m(t)] = o(t^{-1}), \tag{44}$$

where

$$b = \frac{1}{2} \text{tr}(G^T H G).$$

The proof of this theorem is given in Appendix D of Muñoz (1991). As we see from Theorem 3, for a run length t large enough, if $b \neq 0$, we have

$$|\text{Bias}[\alpha'_m(t)]| < |\text{Bias}\{f[r(t)]\}| < |\text{Bias}[\bar{f}_m(t)]|.$$

4.2. Asymptotic Mean Squared Error Expansions

In this subsection, we give asymptotic expansions for the mean squared error of each of the point estimators proposed in §3.1. To provide meaningful results, we shall assume that the stochastic process Y is a strictly stationary stochastic process for which asymptotic independence holds, that is, if $t_2 > t_1$, $Y(t_1)$ and $Y(t_2)$ become statistically independent as $t_2 - t_1 \rightarrow \infty$. The asymptotic independence is expressed in terms of a mixing assumption (see Theorem 4). To provide a precise statement of our main result, we first introduce the appropriate notation.

If $E[\|Y(0)\|^2] < \infty$, we can define the autocovariance function

$$c(u) = E[(Y(t) - r)(Y(t + u) - r)^T], \tag{45}$$

$$t \geq 0, u \geq 0.$$

Note that our steady-state mean r now becomes $r = E[Y(0)]$ and that $c(u)$ is not in general symmetric; however, if we take

$$c^*(u) = \frac{c(u) + [c(u)]^T}{2}, \quad u \geq 0, \tag{46}$$

we can easily verify that $c^*(u)$ is symmetric for any $u \geq 0$. A symmetric matrix is appropriate to describe our results, since it has desirable properties, as for example, the square root $A^{1/2}$ of a matrix A is well defined only if A is positive semidefinite and symmetric. Under suitable mixing assumptions (as those of Theorem 4 below) we can define

$$\Sigma = \int_0^\infty c^*(u) du \quad \text{and} \quad \Sigma_1 = \int_0^\infty uc^*(u) du. \tag{47}$$

Also, provided $E[\|Y(0)\|^3] < \infty$, we consider

$$\mu_3(s_1, s_2, s_3) = E[(\nabla f(r))^T (Y(s_1) - r) \times (Y(s_2) - r)^T H (Y(s_3) - r)], \tag{48}$$

$s_i \geq 0, i = 1, 2, 3$. Note that if Y is a strictly stationary stochastic process that satisfies Assumption 1, then $\Sigma = G G^T$.

The next theorem provides asymptotic expansions for the mean squared errors of the point estimators proposed in §3.1. We shall assume that Y is a strictly stationary and strongly mixing stochastic process (see Ethier and Kurtz 1986 for a definition) and that $\{t^{p/2}\|r(t) - r\|^p : t \geq 0\}$ is uniformly integrable (see Chung 1974 for a definition).

THEOREM 4. *Let Y be a strictly stationary and strongly mixing stochastic process with mixing coefficients $\{\phi(t) : t \geq 0\}$. Assume that Y satisfies Assumption 1, f is polynomially dominated to degree q ($q \geq 0$), and has (finite) partial derivatives of order 4 in a neighborhood of r . Let the number of batches $m \geq 1$ be fixed, and $p = \max\{8, 2q\}$. If $E[\|Y(0)\|^{20}] < \infty$, and there exist constants $t_0 > 0$ and $\epsilon > 0$ such that $\{t^{p/2}\|r(t) - r\|^p : t \geq t_0\}$ is uniformly integrable, and $\phi(t) = O(t^{-(6+\epsilon)})$, as $t \rightarrow \infty$, then*

$$\begin{aligned} \text{mse}\{f[r(t)]\} &= \frac{2a_1}{t} + \frac{2a_2 + 2(b_1 + b_2 + b_3) + c_1^2 + 2c_2 + 4d_1}{t^2} \\ &\quad + O(t^{-5/2}), \end{aligned} \tag{49}$$

as $t \rightarrow \infty$,

$$\begin{aligned} \text{mse}[\bar{f}_m(t)] &= \frac{2a_1}{t} + \frac{2a_2 + 2m(b_1 + b_2 + b_3) + mc_1^2 + m(m+1)c_2 + 4md_1}{t^2} \\ &\quad + O(t^{-5/2}), \end{aligned} \tag{50}$$

and

$$\begin{aligned} \text{mse}[\alpha_m^j(t)] &= \frac{2a_1}{t} \\ &+ \frac{2(m-1)a_2 + 2mc_2}{(m-1)t^2} + O(t^{-5/2}), \end{aligned} \quad (51)$$

where

$$\begin{aligned} a_1 &= [\nabla f(r)]^T \Sigma \nabla f(r), \\ a_2 &= [\nabla f(r)]^T \Sigma_1 \nabla f(r), \\ b_1 &= \int_0^\infty \int_0^\infty \mu_3(0, u_2, u_3) du_2 du_3, \\ b_2 &= \int_0^\infty \int_0^\infty \mu_3(u_1, 0, u_3) du_1 du_3, \\ b_3 &= \int_0^\infty \int_0^\infty \mu_3(u_1, u_2, 0) du_1 du_2, \\ c_1 &= \text{tr}(\Sigma^{1/2} H \Sigma^{1/2}), \\ c_2 &= \sum_{i=1}^d \sum_{j=1}^d q_{ij}^2, \\ d_1 &= \sum_{i=1}^d p_i \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \lambda_{ijkl} \sigma_{ij} \sigma_{kl}, \end{aligned}$$

$\nabla f(r) = (p_1, p_2, \dots, p_d)^T$, $H = (h_{ij})$ and λ_{ijkl} , are defined by

$$p_i = \left. \frac{\partial f(x)}{\partial x_i} \right|_{x=r}$$

for $i = 1, 2, \dots, d$,

$$h_{ij} = \left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=r}$$

for $i, j = 1, 2, \dots, d$,

$$\lambda_{ijkl} = \left. \frac{\partial^4 f(x)}{\partial x_i \partial x_j \partial x_k \partial x_l} \right|_{x=r}$$

for $j, k, l = 1, 2, \dots, d$, q_{ij} and σ_{ij} are the $(i, j)^{\text{th}}$ entries of $Q = \Sigma^{1/2} H \Sigma^{1/2}$ and Σ , respectively.

The proof of this theorem is given in Appendix E of Muñoz (1991). In practice, the assumptions stated in Theorem 4 are quite mild. For example, it is typical that $\|Y(0)\|$ would have finite moments of all orders in realistic discrete-event simulations. In any case, as we can see from Theorem 4, the rates of convergence of the

mean squared errors for the three point estimators that we are considering are the same. Furthermore, we see that

$$\lim_{t \rightarrow \infty} t \text{mse}[\hat{f}_m(t)] = 2a_1,$$

where $\hat{f}_m(t)$ can be any of the three point estimators. This result suggests that all three point estimators exhibit similar performance from the point of view of their mean squared errors. A better asymptotic performance for a particular point estimator cannot be established in general from Theorem 4, since the signs of b_i , $i = 1, 2, 3$ and d_1 will depend not only on the moments of the underlying stochastic process Y but also on the particular function f that we are considering. However, we point out that the coefficient in t^{-2} for the mean squared error expansion of the jackknife estimator is simpler than that of the batch means or the classical estimator and can be smaller under appropriate assumptions.

5. Experimental Results

In this section, we present experimental results obtained from applying our proposed methodology to the estimation of a nonlinear function of a steady-state mean. The system selected to perform our experiments is an $M/M/1$ queue, and we consider the estimation of the variance of the steady-state distribution of the sojourn times (waiting plus service times). To compare the performances of our different point and variability estimators, we compute the (empirical) coverage, the (empirical) bias, the (empirical) mean squared error, the average halfwidth and its standard deviation for all six combinations of point and variability estimators from a number of independent replications. Different numbers of batches were also considered.

The output of our simulation is regarded as a discrete time stochastic process $Z = \{Z_k : k \geq 1\}$, where Z_k denotes the sojourn time of the k th customer, $k = 1, 2, \dots$. As is clear from (8)–(10), the only information required to compute $f[r(t)]$, $\bar{f}_m(t)$ and $\alpha_m^j(t)$ are the values of the batch means $\bar{X}_i(t)$, $i = 1, 2, \dots, m$. Since the steady-state sojourn time distribution in an $M/M/1$ queue is exponential with parameter $\mu(1 - \rho)$, our true steady-state variance becomes

Table 1 Performance of 90% Confidence Regions for the Variance of the Steady-State Sojourn Time Distribution from an M/M/1 Queue Based on 2,000 Independent Replications ($t = \text{run length} = 150,000, \rho = 0.8$)

m	Point Estimator	Bias	mse
5	Classical	-265×10^{-4}	356×10^{-2}
	Batch means	-787×10^{-4}	353×10^{-2}
	Jackknife	-134×10^{-4}	357×10^{-2}
10	Classical	-265×10^{-4}	356×10^{-2}
	Batch means	-1432×10^{-4}	349×10^{-2}
	Jackknife	-135×10^{-4}	357×10^{-2}
20	Classical	-265×10^{-4}	356×10^{-2}
	Batch means	-2761×10^{-4}	344×10^{-2}
	Jackknife	-134×10^{-4}	257×10^{-2}

m	Variability Estimator	Average Halfwidth	Halfwidth St. Deviation
5	Classical	379×10^{-2}	174×10^{-2}
	Jackknife	380×10^{-2}	176×10^{-2}
10	Classical	332×10^{-2}	126×10^{-2}
	Jackknife	336×10^{-2}	131×10^{-2}
20	Classical	312×10^{-2}	106×10^{-2}
	Jackknife	320×10^{-2}	113×10^{-2}

Empirical Coverage (Point Estimator * Variability Estimator)

m	Clas * Bat	Clas * Jack	Bat * Bat	Bat * Jack	Jack * Bat	Jack * Jack
5	0.889	0.889	0.887	0.887	0.890	0.889
10	0.885	0.886	0.883	0.883	0.886	0.886
20	0.882	0.887	0.871	0.876	0.882	0.887

$$\text{Var}[X] = \frac{1}{\mu^2(1 - \rho)^2},$$

where ρ is the traffic intensity. To remove the initial transient effects, Z_1 was sampled from the steady-state distribution.

In Table 1 we summarize the results of 2,000 independent replications with a run length of $t = 150,000$ customers and $\rho = 0.8$. The 2,000 replications ensure a 95% standard error of 0.0109 in the observed coverage. In Tables 2 and 3, we present the results with smaller run lengths ($t = 25,000$ and $t = 75,000$, respectively).

From Table 1 we can see that, if the run length is large enough, all 6 combinations of point and variability estimators provide good coverage, as stated in Theorems 1 and 2. In this case, as we expected, the bias of the jackknife point estimator is smaller (by a half) than the bias of the classical estimator. Also, the largest bias is that of the batch means point estimator. This bias tends to be relatively larger as the number of batches m increases. This result is explained by the fact that the coefficient in t^{-1} in the asymptotic expansion for $\text{Bias}[\bar{f}_m(t)]$ is m times the corresponding coefficient in the asymptotic expansion for $\text{Bias}\{f[r(t)]\}$. A second

Table 2 Performance of 90% Confidence Regions for the Variance of the Steady-State Sojourn Time Distribution from an $M/M/1$ Queue Based on 2,000 Independent Replications (t = run length = 25,000, ρ = 0.8)

m	Point Estimator	Bias	mse
5	Classical	-222×10^{-3}	218×10^{-1}
	Batch means	-527×10^{-3}	205×10^{-1}
	Jackknife	-145×10^{-3}	221×10^{-1}
10	Classical	-222×10^{-3}	218×10^{-1}
	Batch means	-910×10^{-3}	195×10^{-1}
	Jackknife	-145×10^{-3}	221×10^{-1}
20	Classical	-222×10^{-3}	218×10^{-1}
	Batch means	-1665×10^{-3}	184×10^{-1}
	Jackknife	-146×10^{-3}	221×10^{-1}

m	Variability Estimator	Average Halfwidth	Halfwidth St. Deviation
5	Classical	796×10^{-2}	512×10^{-2}
	Jackknife	814×10^{-2}	543×10^{-2}
10	Classical	674×10^{-2}	383×10^{-2}
	Jackknife	714×10^{-2}	435×10^{-2}
20	Classical	590×10^{-2}	298×10^{-2}
	Jackknife	670×10^{-2}	391×10^{-2}

Empirical Coverage (Point Estimator * Variability Estimator)

m	Clas * Bat	Clas * Jack	Bat * Bat	Bat * Jack	Jack * Bat	Jack * Jack
5	0.827	0.829	0.819	0.821	0.828	0.829
10	0.804	0.813	0.784	0.793	0.806	0.816
20	0.771	0.802	0.728	0.748	0.774	0.805

observation from the results of Table 1 is that all three point estimators performed very closely in terms of their mean squared errors. This result agreed with our previous asymptotic expansions of §4.2, since the mean squared error expansions for the three point estimators differ only in the coefficients in t^{-2} , with no evidence that, in general one of these coefficients has to be smaller than another. In general terms, we can see from the results of Table 1 that the jackknife point estimator provided a smaller bias, with no significant increase in the mean squared error with respect to the classical or batch means point estimators.

From Tables 2 and 3, we can see that our results cannot be generalized to the small sample context. In par-

ticular, the bias of the jackknife point estimator can be larger than the bias of the classical estimator (as in Table 3) if the run length is not large enough to provide good coverage. However, all three experiments still have some common patterns. For example, the biases of both the jackknife and the classical point estimators seem to be almost independent of the number of batches m , whereas the bias of the batch means estimators always increases with the number of batches. This result can be explained from the asymptotic expansions of Theorem 3 (note that the constant b in Theorem 3 does not depend on m). With respect to the performance of the variability estimators, we can see from Tables 1, 2, and 3 that the jackknife variability estimator tends to be larger than

Table 3 Performance of 90% Confidence Regions for the Variance of the Steady-State Sojourn Time Distribution from an M/M/1 Queue Based on 2,000 Independent Replications ($t = \text{run length} = 75,000, \rho = 0.8$)

m	Point Estimator	Bias	mse
5	Classical	726×10^{-4}	787×10^{-2}
	Batch means	-313×10^{-4}	766×10^{-2}
	Jackknife	986×10^{-4}	792×10^{-2}
10	Classical	726×10^{-4}	787×10^{-2}
	Batch means	-1633×10^{-4}	748×10^{-2}
	Jackknife	989×10^{-4}	792×10^{-2}
20	Classical	726×10^{-4}	787×10^{-2}
	Batch means	-4315×10^{-4}	717×10^{-2}
	Jackknife	992×10^{-4}	792×10^{-2}

m	Variability Estimator	Average Halfwidth	Halfwidth St. Deviation
5	Classical	523×10^{-2}	292×10^{-2}
	Jackknife	527×10^{-2}	300×10^{-2}
10	Classical	454×10^{-2}	218×10^{-2}
	Jackknife	463×10^{-2}	230×10^{-2}
20	Classical	420×10^{-2}	184×10^{-2}
	Jackknife	440×10^{-2}	206×10^{-2}

Empirical Coverage (Point Estimator * Variability Estimator)

m	Clas * Bat	Clas * Jack	Bat * Bat	Bat * Jack	Jack * Bat	Jack * Jack
5	0.877	0.880	0.875	0.875	0.878	0.880
10	0.858	0.859	0.855	0.857	0.857	0.860
20	0.852	0.867	0.826	0.840	0.854	0.867

the batch means variability estimator, so that it tends to provide better coverage (although the differences are not significant). It is our feeling that this result may be related to the Efron-Stein inequality. Also, we see that in all three experiments a better coverage is obtained either with the jackknife or the classical point estimator combined with the jackknife variability estimator.

6. Conclusions and Recommendations

In this article, we studied the estimation of a nonlinear function $f(r)$ of a multivariate steady-state mean r using

the output of a simulation. We propose a batch-means-based methodology that allows us to consider three point estimators: the classical estimator, the batch means estimator, and the jackknife estimator. We also consider two variability estimators: the batch means variability estimator and the jackknife variability estimator.

Under the assumption that the stochastic process representing the output of the simulation satisfies a functional central limit theorem, we show that all six combinations of the three point estimators and the two variability estimators give asymptotically valid confidence intervals for $f(r)$. To compare the long-run performance

of the three point estimators considered, we obtained asymptotic expansions for the bias and the mean squared errors of the point estimators. Our results show that, for a large run length, the jackknife point estimator has a smaller bias than the classical estimator, and the batch means estimator has the largest bias. In addition, all three point estimators exhibit similar mean squared errors. We run some experiments that confirm these results. However, these experiments show that our asymptotic results can not be extended to the small sample context.

We point out an interesting direction for future research: the development of a general methodology based on the batch means method to estimate other long-run performance measures, such as quantiles of the steady-state distribution (see Muñoz 1991).¹

¹ We thank Jim Wilson, the Associate Editor, and the referee for helpful comments.

References

- Billingsley, P., *Probability and Measure*, 2nd ed., John Wiley, New York, 1986.
- Chung, K. L., *A Course in Probability Theory*, 2nd ed., Academic Press, San Diego, CA, 1974.
- Efron, B. and C. Stein, "The Jackknife Estimate of Variance," *Ann. Statist.*, 9 (1981), 586–596.
- Ethier, S. N. and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley, New York, 1986.
- Fishman, G. S. and L. R. Moore, "Estimating the Mean of a Correlated Binary Sequence with an Application to Discrete-Event Simulation," *J. Assoc. Computing Machinery*, 26 (1979), 82–94.
- Glynn, P. W., "A GSMP Formalism for Discrete-Event Systems," *Proc. IEEE.*, 77 (1989), 14–23.
- and P. Heidelberger, "Jackknifing Under a Budget Constraint," *ORSA J. Computing*, 4 (1992), 226–234.
- and D. L. Iglehart, "Simulation Methods for Queues: An Overview," *Questa*, 3 (1988), 221–255.
- and —, "Simulation Output Analysis Using Standardized Time Series," *Math. Oper. Res.*, 15 (1990), 1–16.
- Iglehart, D. L., "Simulating Stable Stochastic Systems: VI. Quantile Estimation," *J. Assoc. Computing Machinery*, 23 (1976), 347–360.
- Karlin, S. and Y. Rinott, "Applications of ANOVA Type Decompositions for Comparisons of Conditional Variance Statistics Including Jackknife Estimates," *Ann. Statistics*, 10 (1982), 485–501.
- Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis*, 2nd ed., McGraw-Hill, New York, 1991.
- Miller, R. G., "A Trustworthy Jackknife," *Ann. Math. Statistics*, 34 (1964), 1594–1605.
- , "The Jackknife—A Review," *Biometrika*, 61 (1974), 1–15.
- Muñoz, D. F., *Cancellation Methods in the Analysis of Simulation Output*, Ph.D. dissertation, Department of Operations Research, Stanford University, Stanford, CA, 1991.
- Quenouille, M. H., "Notes on Bias in Estimation," *Biometrika*, 43 (1956), 353–360.
- Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, John Wiley, New York, 1980.
- Shedler, G. S., *Regeneration and Networks of Queues*, Springer Verlag, New York, 1987.
- Wilson, J. R. and A. A. B. Pristker, "Evaluation of Startup Policies in Simulation Experiments," *Simulation*, 31 (1978), 79–89.

Accepted by James R. Wilson, former Departmental Editor; received August 1, 1991. This paper has been with the authors 29 months for 4 revisions.