Bias Properties of Budget Constrained Simulations
Author(s): Peter W. Glynn and Philip Heidelberger
Source: *Operations Research*, Vol. 38, No. 5 (Sep. - Oct., 1990), pp. 801-814
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/171038
Accessed: 21/07/2010 03:39

BIAS PROPERTIES OF BUDGET CONSTRAINED SIMULATIONS

PETER W. GLYNN

Stanford University, Stanford, California

PHILIP HEIDELBERGER

IBM T. J. Watson Research Center, Yorktown Heights, New York

(Received March 1988; revision received April 1989; accepted July 1989)

This paper addresses the bias characteristics of estimators produced from Monte Carlo simulations. If the computer time allocated to the simulation is t, then $N(t)$, the number of replications completed by time t, is a renewal process. The simulation implications of a known exact expression for the expected value of a sample mean based on $N(t)$ replications are explored and a similar exact expression for a sample mean based on $N(t) + 1$ replications is derived. Bias expansions for a sample mean based on $N(t)$ or $N(t) + 1$ replications are obtained. The bias in the sample mean based on $N(t)$ replications is at most of order $o(1/t)$. Under suitable moment conditions, the bias decreases at a much faster rate than $o(1/t)$; on the other hand, the estimator based on $N(t) + 1$ replications has bias of order $1/t$. The exact expressions also lead to simple and totally unbiased estimators. Using Taylor series, the bias expansions of a general function of means based on $N(t)$ or $N(t) + 1$ replications are determined. The leading terms in these expansions are of order $1/t$, although the coefficients are different. Based on these expansions, a Tin-style adjusted estimator is proposed to reduce the bias. These expansions are specialized to the case of ratio estimation in regenerative simulation. Due to a cancellation effect, the ratio estimator based on $N(t) + 1$ cycles is biased only to order $o(1/t)$ providing confirmation and reinterpretation of a result of M. Meketon and P. Heidelberger.

This paper is concerned with the problems of analyzing and producing low bias estimates from Monte Carlo simulations. Initially, we consider the problem of estimating $\mu = E[X]$ given a budget constraint t that represents the maximum amount of computer time to be used. This estimation problem arises, for example, in the analysis of transient simulations, in which case X is a performance measure (i.e., real-valued random variable) that depends on simulating the stochastic system under study up to some stopping time for the process. Given the budget constraint, the number of replications $N(t)$ completed by time t will be a random variable. As a consequence, the sample mean $\bar{X}_{N(t)}$ based on the $N(t)$ observations available at time t is a ratio estimator which, in general, is biased; the study of this and related bias issues is the main topic of this paper.

This bias presents a serious potential problem to the simulation analyst in two different applications settings. First, if the time required to generate independent and identically distributed (i.i.d.) copies of X is large (relative to t), then $N(t)$ will be small and the bias could seriously degrade the performance of the estimator $\bar{X}_{N(t)}$. We believe this to be particularly relevant to simulations of large-scale systems, such as those arising in manufacturing, computer and communications systems, which are typically modeled by networks of queues. As these systems, and hence their corresponding simulation models, increase in complexity, we expect the computational requirements of such Monte Carlo simulations to become ever greater.

A second motivation for studying the bias problem concerns estimation of transient performance measures by running independent replications of the simulation, in parallel, on a multiple processor computer. As Heidelberger (1988) has shown, the effects of any bias get magnified and, unless one is very careful, convergence to the wrong quantity can occur. In contrast to the single processor computing environment just discussed, the bias issue is relevant here even if the time required to generate a copy of X is small. A full description of the parallel simulation implications of the results derived here will be treated elsewhere.

Note that alternative stopping rules are also possible. For example, one could run the simulation for a fixed number of replications. In this case, there is no bias

Subject classifications: Probability: regenerative and renewal processes. Simulation: statistical analysis.

0030-364X/90/3805-0801 $01.25

problem in estimating a simple mean, but the completion time is a random variable. Alternatively, sequential stopping procedures could be applied (see, e.g., Lavenberg and Sauer 1975) in which one runs until an accuracy criterion is satisfied. In this case, both the number of replications and the completion time are random variables.

1. OVERVIEW OF RESULTS

In this paper, we will show that the bias of $\bar{X}_{N(t)}$ is surprisingly small (at most $o(1/t)$). We also consider stopping rules that allow the simulation to extend somewhat beyond time t. For example, if one always completes the replication in progress at time t, then resulting estimate, $\bar{X}_{N(t)+1}$, has relatively high bias (of order $1/t$). However, if one extends the simulation beyond time t only if no replications have yet been completed (if $N(t) = 0$), and otherwise stops at time t, then a totally unbiased estimate can be formed. This unbiased estimate is $\bar{X}_{\max(N(t),1)}$, i.e., it is $\bar{X}_{N(t)}$ if $N(t) > 0$ and X_1 if $N(t) = 0$ (in which case, the simulation must be extended beyond time t until the time that the replication completes).

We will more formally describe the mathematical setup and overview the main results of the paper. Let $\tau_k > 0$ be the amount of computer time required to generate X_k, the kth copy of the random variable X. We make the (reasonable) assumption that the sequence of pairs $\{(X_k, \tau_k), k \geq 1\}$ is i.i.d. The number of observations $N(t)$ completed under the budget constraint t constitutes a renewal process with interevent times τ_k (see, e.g., Smith 1955, 1958, 1959, Cox 1962, Karlin and Taylor 1975). The goal of the simulation is to estimate $\mu = E[X_k]$. Note that while we think of τ_k as computer time and t as a time constraint, other interpretations for these variables are possible. For example, t could denote the CPU budget (in dollars) and τ_k could denote the CPU cost of an observation. In the regenerative simulation setting (see, e.g., Crane and Iglehart 1975), t could be the total length (in simulated time) of the run and τ_k could correspond to the length (again in simulated time) of the kth regenerative cycle.

Define $\bar{X}_0 \equiv 0$ and for any $n \geq 1$ let

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

In a simulation of length t, the obvious estimate of $E[X_1]$ is $\bar{X}_{N(t)}$.

Moment expansions are available for both the numerator

$$S_{N(t)} = \sum_{k=1}^{N(t)} X_k$$

and the denominator, $N(t)$, of $\bar{X}_{N(t)}$ (see Smith 1955, 1958, 1959, Brown and Solomon 1975 and Murthy 1974). Under certain technical assumptions, these expressions state that

$$E[S_{N(t)}] = \lambda t E[X_1] + c_1 + o(1)$$

and

$$E[N(t)] = \lambda t + c_2 + o(1) \quad \text{as } t \to \infty$$

where $\lambda = 1/E[\tau_1]$. This suggests that $E[\bar{X}_{N(t)}] = E[X_1] + c/t + o(1/t)$ with $c \neq 0$. However, we show in Section 2 that, under minimal moment assumptions, the bias of $\bar{X}_{N(t)}$ is at most $o(1/t)$. This result is based on a simple expression for $E[\bar{X}_{N(t)}]$

$$E[\bar{X}_{N(t)}] = E[X_1; \tau_1 \leq t] = E[X_1] - E[X_1; \tau_1 > t] \tag{1}$$

where, for a real-valued random variable, Y, $E[Y; A]$ denotes $E[YI(A)]$ and $I(A)$ denotes the indicator of an arbitrary event A. Equation 1 has previously appeared in several references, although it does not seem to be well known (indeed we became aware of these references only after deriving the results of this paper ourselves). Pathak (1976) and Kremers (1986) derive a similar result in the context of survey sampling from a finite population. Kremers states without proof that the result also holds in the so-called infinite population case, which corresponds to the probabilistic setting here. In the survey sampling setting, τ_k could correspond to the cost of performing a sample and t could be the total amount of money available for the survey. Ross (1983, p. 94) contains this result for the special case that $X_k = \tau_k$.

From Equation 1, it may be concluded that

1. Under suitable moment conditions, $|E[\bar{X}_{N(t)}] - E[X_1]|$ goes to zero at a rate much faster than $o(1/t)$. For example, if $E[|X_1 e^{\alpha\tau_1}|] < \infty$ for some $\alpha > 0$, then $|E[\bar{X}_{N(t)}] - E[X_1]| = o(e^{-\alpha t})$.
2. If the τ_k's are uniformly bounded, that is, if there exists a constant B such that $\tau_k \leq B$ a.s., and if $t \geq B$, then $E[\bar{X}_{N(t)}] = E[X_1]$ exactly.
3. If $X_k \geq 0$ a.s., then $E[\bar{X}_{N(t)}] \leq E[X_1]$ and $E[\bar{X}_{N(t)}]$ increases monotonically to $E[X_1]$. This is true no matter what correlation structure exists between X_k and τ_k.
4. An unbiased estimator for $E[X_1]$, $\tilde{X}_{N(t)}$, can always

be formed by setting $\tilde{X}_{N(t)} = \bar{X}_{N(t)}$ if $N(t) > 0$ and $\tilde{X}_{N(t)} = X_1$ if $N(t) = 0$ (note that $\tilde{X}_{N(t)} = \bar{X}_{\max(1, N(t))}$).

The unbiased estimator $\tilde{X}_{N(t)}$ described shows that $\mu = E[X_1]$ can be estimated without bias, provided the simulator agrees to complete the observation in progress at time t if no observations have yet been completed.

In Section 3, we turn to bias theory for $\bar{X}_{N(t)+1}$. Here, Wald's equality (see, e.g., p. 186 of Karlin and Taylor) states that, since $N(t) + 1$ is a stopping time, $E[S_{N(t)+1}] = E[N(t) + 1]E[X_1]$. Given that $E[N(t) + 1]$ is the expected value of the denominator of $\bar{X}_{N(t)+1}$, one might expect that $E[\bar{X}_{N(t)+1}] = E[X_1] + o(1/t)$ as $t \to \infty$. This type of reasoning is valid in the steady-state regenerative simulation setting in which Meketon and Heidelberger (1982) showed that by completing the regenerative cycle in progress at time t and averaging over the resulting $N(t) + 1$ cycles, the bias of the regenerative ratio estimator is reduced from c/t (for $N(t)$ cycles) to $o(1/t)$. In this setting, the denominator of the ratio estimator is the sum (to either $N(t)$ or $N(t) + 1$) of the τ_k's rather than the number of cycles completed. However, this reasoning fails in the current setting. In particular we show that

$$E[\bar{X}_{N(t)+1}] = E[X_1] + E\left[\frac{X_{N(t)+1} - X_1}{N(t) + 1}\right] \tag{2}$$

from which it may be shown that $E[\bar{X}_{N(t)+1}] = E[X_1] + c/t + o(1/t)$ where the constant $c = \mathrm{Cov}[X_k, \tau_k]$. Based on Equation 2, $\bar{X}_{N(t)+1}$ may also be adjusted to yield an unbiased estimator $\tilde{X}_{N(t)+1}$ for $E[X_1]$ which, interestingly, does not involve $X_{N(t)+1}$ at all unless $N(t) = 0$. Furthermore, $\mathrm{Var}[\tilde{X}_{N(t)}] \leqslant \mathrm{Var}[\tilde{X}_{N(t)+1}]$. For the finite sample case, Kremers shows that a different, more complicated, estimator is UMVUE (uniform minimum variance unbiased estimator) for $E[X_1]$ and he states, without conditions or proof, that this estimator is also UMVUE in the infinite population setting. Because Kremers' estimator always requires completing the $(N(t) + 1)$st replication as well as storing and postprocessing $\{(X_k, \tau_k), 1 \leqslant k \leqslant N(t) + 1\}$, we shall not further consider it here (Kremers also recommends $\tilde{X}_{N(t)}$ for sample surveys based on its simplicity).

In Section 4 we consider estimation of a nonlinear function of means. Associated now with replication (cycle) k is a d-dimensional random vector $\mathbf{X}_k = (X_k(1), \ldots, X_k(d))$. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ where $\mu_i = E[X_k(i)]$. The goal of the simulation is now to estimate $g(\boldsymbol{\mu})$ where g is a mapping from R^d to R^1. For example, if $X_k(1)$ is the sum of the response times at a particular queue in the kth regenerative cycle and if $X_k(2)$ is the number of customers served at that queue during the cycle, then $g(\mu_1, \mu_2) = \mu_1/\mu_2$ is the stationary mean response time at the queue.

Even in the case when the number of cycles is a constant N, if the function g is nonlinear and if

$$\bar{\mathbf{X}}_N = (1/N) \sum_{k=1}^{N} \mathbf{X}_k$$

then $g(\bar{\mathbf{X}}_N)$ is a biased estimate of $g(\boldsymbol{\mu})$. In this case, the bias expansion is typically of the form $E[g(\bar{\mathbf{X}}_N)] = g(\boldsymbol{\mu}) + c/N + o(1/N)$ for some constant c (see, e.g., p. 354 of Cramér 1946 or Section 2.8 of Lehmann 1983). A variety of techniques are available to reduce the bias, most notably jackknifing (see Miller 1974 for a survey on the jackknife) or corrections based on Taylor series expansions of the function g (see Beale 1962 and Tin 1965). These techniques, as typically applied, eliminate the $1/N$ term in the bias expansion although higher order methods are available. Iglehart (1975) discusses these and other estimators in the context of ratio estimation in regenerative simulation.

Using Taylor series arguments, central limit theorems for $g(\bar{\mathbf{X}}_{N(t)})$, $g(\tilde{\mathbf{X}}_{N(t)})$ and $g(\bar{\mathbf{X}}_{N(t)+1})$ are stated and their bias expansions derived. In particular, if

$$C_{ij} = \mathrm{Cov}[X_k(i), X_k(j)]$$

and

$$G_{ij} = \left.\frac{\partial^2 g}{\partial x_i \partial x_j}\right|_{\mathbf{x}=\boldsymbol{\mu}}$$

then

$$E[g(\bar{\mathbf{X}}_{N(t)})] = g(\boldsymbol{\mu}) + \frac{E[\tau_k]}{2t} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij} C_{ij} + o\left(\frac{1}{t}\right) \tag{3}$$

(the above expansion is also valid for $E[g(\tilde{\mathbf{X}}_{N(t)})]$). To order $1/t$, this bias expansion is the same as would be obtained if the number of cycles were a constant equal to $t/E[\tau_k]$. For the case of $g(\bar{\mathbf{X}}_{N(t)+1})$ if

$$g_i = \left.\frac{\partial g}{\partial x_i}\right|_{\mathbf{x}=\boldsymbol{\mu}}$$

and $c_i = \mathrm{Cov}[X_k(i), \tau_k]$ (the term of order $1/t$ in the bias expansion of $\bar{X}_{N(t)+1}(i)$), then

$$\begin{aligned}&\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)+1})]\\ &= g(\boldsymbol{\mu}) + \frac{1}{t}\sum_{i=1}^{d} g_i c_i\\ &\quad + \frac{\mathrm{E}[\tau_k]}{2t}\sum_{i=1}^{d}\sum_{j=1}^{d} G_{ij}C_{ij} + o\left(\frac{1}{t}\right). \qquad (4)\end{aligned}$$

There are two bias terms of order $1/t$ in Equation 4. The first term is due to the bias in the sample mean $\bar{\mathbf{X}}_{N(t)+1}$ that arises because the number of observations is random, while the second term is due to the nonlinearity of the function g. In Equation 3, the only bias of order $1/t$ in $\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})]$ and $\mathrm{E}[g(\tilde{\mathbf{X}}_{N(t)})]$ is due to the nonlinearity of the function g. A Tin-style adjustment based on these Taylor series is proposed as a bias reducing technique.

In Section 5, the case of ratio estimators is reexamined and the results of Meketon and Heidelberger are verified and reinterpreted. As mentioned, Meketon and Heidelberger showed, using different methods, that if $g(\mu_1, \mu_2) = \mu_1/\mu_2$ and $X_k(2) = \tau_k$, then, under suitable technical conditions, $\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})] = \mu_1/\mu_2 + c/t + O(1/t^2)$ where the constant c was explicitly calculated and $\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)+1})] = \mu_1/\mu_2 + O(1/t^2)$. This case corresponds to the steady-state estimation problem in regenerative simulation for which t is measured in units of simulated time, as opposed to *real* computer time. This reduction in bias was seen to come about as a result of Wald's Equation, the fact that, because t is simulated time, the denominator in the ratio cannot differ greatly from t and, not so obviously, the particular form of the function g. When viewed as a special case of the results of Section 4, we show that the two terms of order $1/t$ in Equation 4 for this particular function g (perhaps luckily) cancel each other out, resulting in the bias reduction. Thus, the order of magnitude advantage that the ratio estimator enjoys by completing the cycle in progress at time t disappears for general functions g. Glynn (1987) discusses an alternative bias reducing technique for this case as well. Glynn's method does not require completion of the cycle in progress at time t, although we emphasize that in that setting t must also be equal to, or proportional to, simulated time rather than computer time. Finally, Section 6 summarizes the results of this paper.

2. THE SAMPLE MEAN WITH $N(t)$ OBSERVATIONS

In this section we consider $\bar{X}_{N(t)}$. Throughout, we assume that the pairs $\{(X_k, \tau_k), k \geq 1\}$ are i.i.d. and that $0 < \tau_k < \infty$ a.s. We let $\Rightarrow$ denote weak convergence or convergence in distribution (Billingsley 1968).

The results of this paper depend upon the observation that given $N(t) \geq 1$, the pairs of random variables $(X_1, \tau_1), \ldots, (X_{N(t)}, \tau_{N(t)})$ are exchangeable (see p. 349 of Chung 1974, Pathak, p. 94 of Ross or Kremers); we call this *conditional exchangeability*. To see this, let $\mathbf{Z}_k = (X_k, \tau_k)$ and let $\sigma(1), \ldots, \sigma(k)$ be any permutation of the integers $1, \ldots, k$. Then

$$\begin{aligned}&\mathbf{P}\{(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) \in \mathbf{A} \mid N(t) = k\}\\ &= \frac{\mathbf{P}\{(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) \in \mathbf{A};\ \tau_1 + \ldots + \tau_k \leq t < \tau_1 + \ldots + \tau_{k+1}\}}{P\{\tau_1 + \ldots + \tau_k \leq t < \tau_1 + \ldots + \tau_k + \tau_{k+1}\}}\\ &= \frac{\mathbf{P}\{(\mathbf{Z}_{\sigma(1)}, \ldots, \mathbf{Z}_{\sigma(k)}) \in \mathbf{A};\ \tau_{\sigma(1)} + \ldots + \tau_{\sigma(k)} \leq t < \tau_{\sigma(1)} + \ldots + \tau_{\sigma(k)} + \tau_{k+1}\}}{P\{\tau_{\sigma(1)} + \ldots + \tau_{\sigma(k)} \leq t < \tau_{\sigma(1)} + \ldots + \tau_{\sigma(k)} + \tau_{k+1}\}}\\ &= \frac{\mathbf{P}\{(\mathbf{Z}_{\sigma(1)}, \ldots, \mathbf{Z}_{\sigma(k)}) \in \mathbf{A};\ \tau_1 + \ldots + \tau_k \leq t < \tau_1 + \ldots + \tau_k + \tau_{k+1}\}}{P\{\tau_1 + \ldots + \tau_k \leq t < \tau_1 + \ldots + \tau_k + \tau_{k+1}\}}\\ &= \mathbf{P}\{(\mathbf{Z}_{\sigma(1)}, \ldots, \mathbf{Z}_{\sigma(k)}) \in \mathbf{A} \mid N(t) = k\} \qquad (5)\end{aligned}$$

where the second equality follows because $\{(X_i, \tau_i), i \geq 1\}$ are i.i.d. and the third equality relies on $\tau_1 + \ldots + \tau_k = \tau_{\sigma(1)} + \ldots + \tau_{\sigma(k)}$. Using this conditional exchangeability, the exact expression for $\mathrm{E}[\bar{X}_{N(t)}]$ is easy to derive (see p. 94 of Ross): By conditional exchangeability

$$\mathrm{E}[X_j \mid N(t) = k] = \mathrm{E}[X_1 \mid N(t) = k]$$

for each $1 \leq j \leq k$ and therefore

$$\mathrm{E}[\bar{X}_{N(t)} \mid N(t) = k] = \mathrm{E}[X_1 \mid N(t) = k].$$

Multiplying this equation by $\mathbf{P}\{N(t) = k\}$ and summing over $k \geq 1$ yields the following theorem.

Theorem 1. *If* $\mathrm{E}[\,|X_1|\,] < \infty$, *then*

$$\begin{aligned}\mathrm{E}[\bar{X}_{N(t)}] &= \mathrm{E}[X_1;\ \tau_1 \leq t]\\ &= \mathrm{E}[X_1] - \mathrm{E}[X_1;\ \tau_1 > t]. \qquad (6)\end{aligned}$$

While quantities such as $\mathrm{E}[X_1 \mid N(t) = k]$ and $\mathrm{E}[X_1;\ \tau_1 \leq t]$ cannot usually be evaluated analytically, it is a straightforward exercise to verify Theorem 1 by direct calculation in a particular case: computation of

$$\mathrm{E}[\bar{\tau}_{N(t)}] = \mathrm{E}\left[\frac{\sum_{k=1}^{N(t)} \tau_k}{N(t)};\ N(t) \geq 1\right]$$

for a Poisson process (since, for example, in this case $E[\tau_1 \mid N(t) = k] = t/(k + 1)$).

From Theorem 1, we have the following corollaries which establish convergence rates of $E[\bar{X}_{N(t)}]$ to $E[X_1]$ under a variety of moment assumptions on X_k and τ_k. We begin by observing that if the τ_k's are bounded, and if t is larger than the bound, then there is absolutely no sampling bias whatsoever in the estimator $\bar{X}_{N(t)}$, i.e., $E[\bar{X}_{N(t)}] = E[X_1]$ exactly.

Corollary 1. *If* $E[\,|X_1|\,] < \infty$ *and there exists a constant* $B > 0$ *such that* $\tau_k \leqslant B$ *a.s., then for any* $t \geqslant B$, $E[\bar{X}_{N(t)}] = E[X_1]$.

Corollary 2. *If* $E[\,|X_1|\,] < \infty$ *and* $X_k \geqslant 0$ *a.s., then* $E[\bar{X}_{N(t)}] \leqslant E[X_1]$ *and* $E[\bar{X}_{N(t)}]$ *increases monotonically to* $E[X_1]$.

Corollary 3. *If* $E[\,|X_1|\,] < \infty$ *and* $X_k \geqslant 0$ *a.s., then*

$$\int_0^\infty |\,E[\bar{X}_{N(t)}] - E[X_1]\,|\; dt = E[X_1 \tau_1].$$

Proof. By Theorem 1 and Fubini's theorem

$$\int_0^\infty |\,E[\bar{X}_{N(t)}] - E[X_1]\,|\; dt$$

$$= \int_0^\infty E[X_1 ; \tau_1 > t]\; dt$$

$$= E\left[\int_0^\infty X_1 I(\tau_1 > t)\; dt\right] = E[X_1 \tau_1]. \tag{7}$$

Corollary 4. *If* $E[\,|X_1|\,] < \infty$, *then*

$$\int_0^\infty |\,E[\bar{X}_{N(t)}] - E[X_1]\,|\; dt \leqslant E[\,|X_1 \tau_1|\,].$$

Corollary 5. *If* $E[\,|X_1|\,] < \infty$ *and* $E[\,|X_1 \tau_1^p|\,] < \infty$ *for some* $p > 0$, *then*

$$E[\bar{X}_{N(t)}] = E[X_1] + o(t^{-p}).$$

Proof. For $t > 0$, $1 \leqslant \tau_1/t$ on the set $\{\tau_1 > t\}$ and therefore

$$|\,E[\bar{X}_{N(t)}] - E[X_1]\,| \leqslant E[\,|X_1|\,; \tau_1 > t]$$
$$\leqslant \frac{E[\,|X_1 \tau_1^p|\,; \tau_1 > t]}{t^p}. \tag{8}$$

By the dominated convergence theorem, $E[\,|X_1 \tau_1^p|\,; \tau_1 > t] \to 0$ and thus $t^p |\,E[\bar{X}_{N(t)}] - E[X_1]\,| \to 0$.

Thus, under the minimal moment assumptions $E[\,|X_1|\,] < \infty$ and $E[\,|X_1 \tau_1|\,] < \infty$, then $E[\bar{X}_{N(t)}] = E[X_1] + o(1/t)$. A similar proof to Corollary 5 establishes that $E[\bar{X}_{N(t)}]$ can converge exponentially fast to $E[X_1]$ under the appropriate moment condition.

Corollary 6. *If* $E[\,|X_1 e^{\alpha \tau_1}|\,] < \infty$ *for some* $\alpha > 0$, *then* $E[\bar{X}_{N(t)}] = E[X_1] + o(e^{-\alpha t})$.

By Cauchy-Schwarz the conditions of Corollary 5 will be satisfied if $E[X_1^2] < \infty$ and $E[\tau_1^{2p}] < \infty$ and the conditions of Corollary 6 will be satisfied if $E[X_1^2] < \infty$ and $E[e^{2\alpha \tau_1}] < \infty$.

Based on Theorem 1, there is a simple adjustment to $\bar{X}_{N(t)}$ that results in a totally unbiased estimator for $E[X_1]$. Define $\tilde{X}_{N(t)} = \bar{X}_{N(t)}$ if $N(t) > 0$ and $\tilde{X}_{N(t)} = X_1$ if $N(t) = 0$.

Corollary 7. *If* $E[\,|X_1|\,] < \infty$ *then* $E[\tilde{X}_{N(t)}] = E[X_1]$.

Note that if $E[\,|X_1|\,] < \infty$, then

$$\lim_{t\to\infty} \bar{X}_{N(t)} = \lim_{t\to\infty} \tilde{X}_{N(t)} = E[X_1] \quad \text{a.s.}$$

In addition, since

$$\sqrt{t}(\tilde{X}_{N(t)} - \bar{X}_{N(t)}) = \sqrt{t} X_1 I(N(t) = 0) \Rightarrow 0$$

then $\tilde{X}_{N(t)}$ obeys the same well known central limit theorems as $\bar{X}_{N(t)}$ (see, e.g., Theorem 17.1 of Billingsley): if $\sigma^2 = \mathrm{Var}[X_1]$ and $0 < E[\tau_1] < \infty$, then

$$\sqrt{N(t)}(\tilde{X}_{N(t)} - E[X_1]) \Rightarrow \sigma N(0, 1)$$
$$\sqrt{t}(\tilde{X}_{N(t)} - E[X_1]) \Rightarrow \sigma E[\tau_1]^{1/2} N(0, 1) \tag{9}$$

where $N(a, b)$ denotes a normally distributed random variable with mean a and variance b. To apply these central limit theorems for the purpose of generating confidence intervals for $E[X_1]$ requires estimating the standard deviation term σ. Note that if $E[X_1^2] < \infty$, then

$$\lim_{t\to\infty} \hat{\sigma}^2(t) = \sigma^2 \quad \text{a.s.}$$

where we define $\hat{\sigma}^2(t) = 0$ if $N(t) \leqslant 1$ and

$$\hat{\sigma}^2(t) = \sum_{k=1}^{N(t)} (X_k - \bar{X}_{N(t)})^2/(N(t) - 1) \quad \text{if } N(t) \geqslant 2.$$

Thus, the estimator $\hat{\sigma}^2(t)$, in conjunction with the first central limit theorem of Equation 9, can be used to obtain confidence intervals for $E[X_1]$.

The estimator $\tilde{X}_{N(t)}$ only requires completing the replication in progress at time t if there are no observations by time t, i.e., if $N(t) = 0$. Otherwise, if $N(t) > 0$, the simulation stops at exactly time t. Letting T_t denote the completion time, then $T_t = tI(\tau_1 \leqslant t) + \tau_1 I(\tau_1 > t)$ and therefore $\lim_{t\to\infty} T_t/t = 1$ a.s. Furthermore, $E[T_t] = t\mathbf{P}\{\tau_1 \leqslant t\} + E[\tau_1 ; \tau_1 > t]$.

Thus

$$\mathrm{E}[T_t] - t = \mathrm{E}[\tau_1;\, \tau_1 > t] - t\mathbf{P}\{\tau_1 > t\}. \tag{10}$$

A similar argument to that given in Corollary 5 can be used to bound the right-hand side of Equation 10. For example, we get the following corollary.

Corollary 8. *If* $\mathrm{E}[\tau_1^{1+p}] < \infty$ *for some* $p > 0$, *then* $\mathrm{E}[T_t] - t = \mathrm{o}(t^{-p})$.

Proof. By Exercise 15 on p. 49 of Chung $t^{p+1}\mathbf{P}\{\tau_1 > t\} = \mathrm{o}(1)$ and thus $t\mathbf{P}\{\tau_1 > t\} = \mathrm{o}(t^{-p})$. Applying Corollary 5 with $X_1 = \tau_1$ shows that $t^p\mathrm{E}[\tau_1;\, \tau_1 > t] \to 0$.

The fact that the bias in $\bar{X}_{N(t)}$ is at most $\mathrm{o}(1/t)$ can also be established by using a multidimensional central limit theorem for cumulative processes followed by an application of uniform integrability. Indeed, we originally observed and established that $\mathrm{E}[\bar{X}_{N(t)}] = \mathrm{E}[X_1] + \mathrm{o}(1/t)$ by using these arguments. Since this alternative approach is more in keeping with conventional methods for proving bias expansions (see Cramér), and because of the surprising $\mathrm{o}(1/t)$ magnitude of the bias, we offer the following proof. Define $\hat{X}_k = X_k - \mathrm{E}[X_1]$ and $\hat{S}_n = \sum_{k=1}^{n} \hat{X}_k$. Then

$$\bar{X}_{N(t)} - \mathrm{E}[X_1] = -\mathrm{E}[X_1] \quad \text{if } N(t) = 0$$

and

$$\bar{X}_{N(t)} - \mathrm{E}[X_1] = \frac{\hat{S}_{N(t)}}{N(t)} \quad \text{if } N(t) > 0.$$

Thus

$$\begin{aligned}\mathrm{E}[\bar{X}_{N(t)}] &- \mathrm{E}[X_1]\\ &= -\mathrm{E}[X_1]\mathbf{P}\{N(t) = 0\}\\ &\quad + \mathrm{E}[I(N(t) > 0)\hat{S}_{N(t)}/N(t)].\end{aligned} \tag{11}$$

But $\mathbf{P}\{N(t) = 0\} = \mathbf{P}\{\tau_1 > t\} \le \mathrm{E}[\tau_k^2]/t^2 = \mathrm{o}(1/t)$, so the first term in Equation 11 can be ignored. Let $\lambda = 1/\mathrm{E}[\tau_1]$. Since

$$\frac{I(N(t) > 0)}{N(t)} = I(N(t) > 0)\left(\frac{1}{\lambda t} + \frac{1}{N(t)} - \frac{1}{\lambda t}\right)$$

the second term in Equation 11 can be written as $b_1(t) + b_2(t)$ where

$$b_i(t) = \mathrm{E}[B_i(t)]$$

$$B_1(t) = \frac{I(N(t) > 0)\hat{S}_{N(t)}}{\lambda t}$$

and

$$\begin{aligned}B_2(t) &= I(N(t) > 0)\hat{S}_{N(t)}\left(\frac{1}{N(t)} - \frac{1}{\lambda t}\right)\\ &= \frac{I(N(t) > 0)}{\lambda N(t)}\,\frac{\hat{S}_{N(t)}}{\sqrt{t}}\,\frac{\lambda t - N(t)}{\sqrt{t}}.\end{aligned} \tag{12}$$

Because the denominator of $B_1(t)$ is deterministic, existing bias expansions for $\hat{S}_{N(t)}$ can be used to obtain $b_1(t) = -\mathrm{Cov}[X_k, \tau_k]/t + \mathrm{o}(1/t)$ (see for example, Lemma 1 of Brown and Solomon). Using a multivariate central limit theorem and uniform integrability, it will be shown that $b_2(t) = \mathrm{Cov}[X_k, \tau_k]/t + \mathrm{o}(1/t)$ and thus $b_1(t) + b_2(t) = \mathrm{o}(1/t)$. By the multidimensional central limit theorem for cumulative processes (e.g., apply the Cramér–Wold device (p. 49 of Billingsley) to either the central limit theorem in Smith (1958) or Theorem 17.1 of Billingsley)

$$\left(\frac{\hat{S}_{N(t)}}{\sqrt{t}}, \frac{\lambda t - N(t)}{\sqrt{t}}\right) \Rightarrow \mathbf{N}(\mathbf{0}, \lambda\mathbf{A})$$

where $\mathbf{N}(\mathbf{0}, \lambda\mathbf{A}) = (N_1, N_2)$ is a bivariate normal distribution with means 0 and covariance matrix $\lambda\mathbf{A}$ with $A_{11} = \mathrm{Var}[X_k]$, $A_{22} = \mathrm{Var}[\lambda\tau_k]$ and $A_{12} = A_{21} = \lambda\,\mathrm{Cov}[X_k, \tau_k]$. Thus, $tB_2(t) \Rightarrow (1/\lambda^2)N_1N_2$ and, assuming uniform integrability, $tb_2(t) \to \mathrm{Cov}[X_k, \tau_k]$. The uniform integrability of $tB_2(t)$ can be established under suitable moment conditions using the results of Chow, Hsiung and Lai (1979), but higher moments must be assumed finite than those given in Corollary 5). Such uniform integrability will be considered in Section 4.

3. THE SAMPLE MEAN WITH $N(t) + 1$ OBSERVATIONS

In this section, we consider $\bar{X}_{N(t)+1}$. We begin with the analog of Theorem 1, a simple expression for $\mathrm{E}[\bar{X}_{N(t)+1}]$.

Theorem 2. *If* $\mathrm{E}[|X_1|] < \infty$, *then*

$$\begin{aligned}\mathrm{E}[\bar{X}_{N(t)+1}]&\\ = \mathrm{E}[X_1] &+ \mathrm{E}\left[\frac{X_{N(t)+1}}{N(t) + 1}\right] - \mathrm{E}\left[\frac{X_1}{N(t) + 1}\right].\end{aligned} \tag{13}$$

Proof. The proof uses a conditioning argument similar to that used in Theorem 1. Let $S_n = X_1 + \ldots + X_n$ and note that

$$\bar{X}_{N(t)+1} = \frac{S_{N(t)}}{N(t) + 1} + \frac{X_{N(t)+1}}{N(t) + 1}. \tag{14}$$

Then, for any $k \geqslant 0$

$$\mathrm{E}\left[\frac{S_{N(t)}}{N(t)+1}\,\middle|\, N(t)=k\right]$$

$$= \frac{k}{k+1}\,\mathrm{E}[X_1 \mid N(t)+k]$$

$$= \mathrm{E}[X_1 \mid N(t)=k]$$

$$- \mathrm{E}\left[\frac{X_1}{N(t)+1}\,\middle|\, N(t)=k\right] \quad (15)$$

and the result follows by multiplying Equation 15 by $\mathbf{P}\{N(t)=k\}$ and summing over $k \geqslant 0$.

We next consider the uniform integrability of $\bar{X}_{N(t)}$ and $\bar{X}_{N(t)+1}$. These results will be used to obtain the bias expansion of $\mathrm{E}[\bar{X}_{N(t)+1}]$.

Lemma 1. *If* $\mathrm{E}[|X_1|^{1+\delta}] < \infty$ *for some* $\delta > 0$, *then* $\mathrm{E}[\bar{X}_{N(t)}|]$ *and* $\mathrm{E}[\bar{X}_{N(t)+1}|]$ *are bounded functions of t.*

Proof. Consider $\bar{X}_{N(t)+1}$

$$|\bar{X}_{N(t)+1}| \leqslant \frac{\sum_{k=1}^{N(t)+1}|X_k|}{N(t)+1} \leqslant \sup_{n\geqslant 1}\frac{\sum_{k=1}^{n}|X_k|}{n}. \quad (16)$$

By Exercise 11 on p. 356 of Chung, the expected value of the right-hand side of (16) is finite provided that $\mathrm{E}[|X_1| \log^+(|X_1|)] < \infty$ where $\log^+(x) = \max(\log(x), 0)$. But this expectation is finite provided that $\mathrm{E}[|X_1|^{1+\delta}] < \infty$ for some $\delta > 0$. The proof for $\bar{X}_{N(t)}$ is similar.

Corollary 9. *If* $\mathrm{E}[|X_1|^{\beta+\delta}] < \infty$ *for some* $\beta \geqslant 1$ *and* $\delta > 0$, *then* $\mathrm{E}[|\bar{X}_{N(t)}|^{\beta}]$ *and* $\mathrm{E}[|\bar{X}_{N(t)+1}|^{\beta}]$ *are bounded functions of t.*

Proof. By the convexity of $h(x) = x^{\beta}$ for $\beta \geqslant 1$ and $x \geqslant 0$

$$|\bar{X}_{N(t)+1}|^{\beta} \leqslant \frac{\sum_{k=1}^{N(t)+1}|X_k|^{\beta}}{(N(t)+1)}$$

and the result follows directly from Lemma 1 (identify X_k appearing in Lemma 1 with $|X_k|^{\beta}$ here).

Corollary 10. *If* $\mathrm{E}[|X_1|^{\beta+\delta}] < \infty$ *for some* $\beta \geqslant 1$ *and* $\delta > 0$, *then*

$$\lim_{t\to\infty} \mathrm{E}[|\bar{X}_{N(t)}|^{\beta}] = \lim_{t\to\infty} \mathrm{E}[|\bar{X}_{N(t)+1}|^{\beta}]$$

$$= \mathrm{E}[X_1]|^{\beta}. \quad (17)$$

Proof. Since, for example, $|\bar{X}_{N(t)+1}|^{\beta} \Rightarrow |\mathrm{E}[X_1]|^{\beta}$, the result is true provided the family $\{|\bar{X}_{N(t)+1}|^{\beta}\}$ is uniformly integrable which is true provided that $\mathrm{E}[|\bar{X}_{N(t)+1}|^{\beta+\epsilon}]$ is bounded for some $\epsilon > 0$. By Corollary 9, these expectations are bounded if $\epsilon < \delta$.

Corollary 11. *If* $\mathrm{E}[\tau_1^{\beta+\delta}] < \infty$ *for some* $\beta \geqslant 1$ *and* $\delta > 0$, *then* $\mathrm{E}[(t/N(t))^{\beta};\ N(t) > 0]$ *is a bounded function of t.*

Proof. If $N(t) > 0$, then

$$\frac{t}{N(t)} \leqslant \frac{N(t)+1}{N(t)}\,\frac{\sum_{k=1}^{N(t)+1}\tau_k}{N(t)+1} \leqslant 2\bar{\tau}_{N(t)+1} \quad (18)$$

and the result follows from Corollary 9.

Corollary 12. *If* $\mathrm{E}[\tau_1^{\beta+\delta}] < \infty$ *for some* $\beta \geqslant 1$ *and* $\delta > 0$, *then*

$$\lim_{t\to\infty} \mathrm{E}\left[\left(\frac{t}{N(t)}\right)^{\beta};\ N(t) > 0\right] = \mathrm{E}[\tau_1]^{\beta}.$$

We now use the above results to analyze the remainder terms in Theorem 2.

Lemma 2. *If* $\mathrm{E}[\tau_1] < \infty$, *then*

$$\lim_{t\to\infty} \frac{tX_1}{N(t)+1} = \mathrm{E}[\tau_1]X_1 \quad \textit{a.s.} \quad (19)$$

$$\frac{tX_{N(t)+1}}{N(t)+1} \Rightarrow \mathrm{E}[\tau_1]X^* \quad \textit{as } t \to \infty \quad (20)$$

where $\mathbf{P}\{X^* \in A\} = \mathrm{E}[\tau_1;\ X_1 \in A]/\mathrm{E}[\tau_1]$.

Proof. Since $\lim_{t\to\infty} t/(N(t)+1) = \mathrm{E}[\tau_1]$ a.s., the proof is complete provided that $X_{N(t)+1} \Rightarrow X^*$. The proof of this result is standard if τ_1 has a nonlattice distribution (see, e.g., Exercise 3.20 on p. 97 of Ross). For completeness, we provide the entire proof including the lattice case. Let $a(t) = \mathbf{P}\{X_{N(t)+1} \in A\}$. Then, by conditioning on τ_1, $a = b + F{*}a$ where F is the distribution function of τ_1, $*$ denotes convolution and $b(t) = \mathbf{P}\{X_1 \in A;\ \tau_1 > t\}$. Note that $b(t)$ is monotone, nonincreasing and integrable and, hence, directly Riemann integrable. They key renewal theorem then shows that

$$\lim_{t\to\infty} a(t) = \int_0^{\infty} b(t)\,dt \Big/ \mathrm{E}[\tau_1]$$

if τ_1 has a nonlattice distribution. By Fubini's theorem

$$\int_0^{\infty} b(t)\,dt = \mathrm{E}[\tau_1;\ X_1 \in A]$$

completing the argument in the nonlattice case. If τ_1

has a lattice distribution with span h, then

$$\lim_{n\to\infty} a(nh) = \frac{h}{\mathrm{E}[\tau_1]} \sum_{k=0}^{\infty} \mathbf{P}\{\tau_1 > kh;\ X_1 \in A\}$$

$$= \frac{\mathrm{E}[\tau_1;\ X_1 \in A]}{\mathrm{E}[\tau_1]}. \tag{21}$$

Furthermore, $a(t) = a(nh)$ for $nh \leqslant t < (n+1)h$. Thus, even in the lattice case $a(t) \to \mathbf{P}(X^* \in A\}$ and therefore $X_{N(t)+1} \Rightarrow X^*$ as $t \to \infty$.

Lemma 3. *If* $\mathrm{E}[\,|X_1|^{2+\delta}] < \infty$ *and* $\mathrm{E}[\tau_1^{2+\delta}] < \infty$ *for some* $\delta > 0$, *then*

$$\lim_{t\to\infty} \mathrm{E}\left[\frac{tX_1}{N(t)+1}\right] = \mathrm{E}[\tau_1]\mathrm{E}[X_1]. \tag{22}$$

Proof. By Lemma 2, we need only show that $tX_1/(N(t)+1)$ is uniformly integrable. By Cauchy–Schwarz

$$\mathrm{E}\left[\left|\frac{tX_1}{N(t)+1}\right|^{1+\epsilon/2}\right]$$

$$\leqslant \mathrm{E}[\,|X_1|^{2+\epsilon}]^{1/2}\mathrm{E}\left[\left|\frac{t}{N(t)+1}\right|^{2+\epsilon}\right]^{1/2} \tag{23}$$

which is bounded by Corollary 11 provided that $\epsilon < \delta$.

Lemma 4. *If* $\mathrm{E}[\,|X_1|^{2+\delta}\tau_1] < \infty$ *and* $\mathrm{E}[\tau_1^{2+\delta}] < \infty$ *for some* $\delta > 0$, *then*

$$\lim_{t\to\infty} \mathrm{E}\left[\frac{tX_{N(t)+1}}{N(t)+1}\right] = \mathrm{E}[\tau_1]\mathrm{E}[X^*] = \mathrm{E}[X_1\tau_1]. \tag{24}$$

Proof. We need to show that $tX_{N(t)+1}/(N(t)+1)$ is uniformly integrable, which by the same reasoning as in Lemma 3 reduces to showing that $\mathrm{E}[\,|X_{N(t)+1}|^{2+\epsilon}]$ is bounded. However, by expressing $a(t) = \mathrm{E}[\,|X_{N(t)+1}|^{2+\epsilon}]$ as the solution to a renewal equation, we can apply the same argument as in Lemma 2 to conclude that

$$\lim_{t\to\infty} \mathrm{E}[\,|X_{N(t)+1}|^{2+\epsilon}] = \frac{\mathrm{E}[\,|X_1|^{2+\epsilon}\tau_1]}{\mathrm{E}[\tau_1]} < \infty \quad \text{for } \epsilon \leqslant \delta.$$

Combining the above results yields the leading term in the bias expansion of $\bar{X}_{N(t)+1}$.

Theorem 3. *If* $\mathrm{E}[\,|X_1|^{2+\delta}] < \infty$, $\mathrm{E}[\,|X_1|^{2+\delta}\tau_1] < \infty$ *and* $\mathrm{E}[\tau_1^{2+\delta}] < \infty$ *for some* $\delta > 0$, *then*

$$\mathrm{E}[\bar{X}_{N(t)+1}] = \mathrm{E}[X_1] + \frac{\mathrm{Cov}[X_1,\ \tau_1]}{t} + o\left(\frac{1}{t}\right). \tag{25}$$

Proof. Since $\mathrm{Cov}[X_1,\ \tau_1] = \mathrm{E}[X_1,\ \tau_1] - \mathrm{E}[X_1]\mathrm{E}[\tau_1]$, the result follows immediately from Theorem 2 and Lemmas 3 and 4.

Notice that if X_k and τ_k are positively correlated, then $\mathrm{E}[\bar{X}_{N(t)+1}] \geqslant \mathrm{E}[X_1]$ for large values of t. However, the analog of Corollary 2 is not true.

The expansion of $\mathrm{E}[\bar{X}_{N(t)+1}]$ can also be established using Wald's Equation, the central limit theorem and uniform integrability. As in Section 2, write $\mathrm{E}[\bar{X}_{N(t)+1}] - \mathrm{E}[X_1] = \mathrm{E}[C_1(t)] + \mathrm{E}[C_2(t)]$ where $C_1(t) = \hat{S}_{N(t)+1}/(\lambda t)$ and

$$C_2(t) = \frac{1}{\lambda(N(t)+1)} \frac{\hat{S}_{N(t)+1}}{\sqrt{t}} \frac{\lambda t - (N(t)+1)}{\sqrt{t}}. \tag{26}$$

By Wald's Equation $\mathrm{E}[C_1(t)] = 0$. The multidimensional central limit theorem is still valid when $N(t)+1$ replaces $N(t)$. Thus, provided that $tC_2(t)$ is uniformly integrable

$$\mathrm{E}[tC_2(t)] \to \mathrm{Cov}[X_k,\ \tau_k].$$

The necessary uniform integrability can be established using the results of Chow, Hsiung and Lai.

Theorem 2 suggests an unbiased adjustment to $\bar{X}_{N(t)+1}$. Define

$$\tilde{X}_{N(t)+1} = \bar{X}_{N(t)+1} + \frac{X_1 - X_{N(t)+1}}{N(t)+1}.$$

By Theorem 2, if $\mathrm{E}[\,|X_1|\,] < \infty$, then $\mathrm{E}[\tilde{X}_{N(t)+1}] = \mathrm{E}[X_1]$. If $N(t) = 0$, then $\tilde{X}_{N(t)+1} = \tilde{X}_{N(t)} = X_1$ while if $N(t) > 0$, then

$$\tilde{X}_{N(t)+1} = \frac{X_1 + \ldots + X_{N(t)}}{N(t)+1} + \frac{X_1}{N(t)+1}$$

$$= \frac{N(t)}{N(t)+1}\bar{X}_{N(t)} + \frac{X_1}{N(t)+1}. \tag{27}$$

Notice that (27) does not involve $X_{N(t)+1}$, and thus, in contrast to $\bar{X}_{N(t)+1}$, $\tilde{X}_{N(t)+1}$ can be formed at time t provided that $N(t) > 0$. Equation 29 has the following interpretation: by Theorem 2, the term $\bar{X}_{N(t)}N(t)/(N(t)+1)$ is too small, to produce an unbiased estimate, by a factor of $N(t)/(N(t)+1)$. Therefore, it needs to be adjusted by adding in a typical term $X_1/(N(t)+1)$, and X_1 is thus counted twice. Using conditional exchangeability, any X_k for $1 \leqslant k \leqslant N(t)$ could be added as the compensating term. Let $\tilde{X}_{N(t)+1}(k)$ denote the estimator of this form when X_k is the compensating term. Notice that if $N(t) > 0$, then

$$\tilde{X}_{N(t)} = \frac{\sum_{k=1}^{N(t)} \tilde{X}_{N(t)+1}(k)}{N(t)}. \tag{28}$$

Equation 28 and the asymmetry of $\tilde{X}_{N(t)+1}$ suggest that $\tilde{X}_{N(t)}$ is a more efficient estimator for $E[X_1]$ than $\tilde{X}_{N(t)+1}$. The following theorem establishes that this is indeed the case.

Theorem 4. *If* $E[X_1^2] < \infty$, *then* $\text{Var}[\tilde{X}_{N(t)}] \leq \text{Var}[\tilde{X}_{N(t)+1}]$.

Proof. Since $\tilde{X}_{N(t)}$ and $\tilde{X}_{N(t)+1}$ are both unbiased for $E[X_1]$, it suffices to prove that $E[\tilde{X}^2_{N(t)}] \leq E[\tilde{X}^2_{N(t)+1}]$. Write $\tilde{X}^2_{N(t)+1}$ as

$$\tilde{X}^2_{N(t)+1} = \tilde{X}^2_{N(t)} + 2\tilde{X}_{N(t)}(\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)}) + (\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)})^2 \quad (29)$$

and note that each of the terms on the right-hand side of (29) is bounded by a multiple of

$$\left(\sum_{k=1}^{N(t)+1} |X_k|\right)^2.$$

Wald's second moment equality (p. 24 of Chow, Robbins and Siegmund 1971), can be applied, since $E[N(t)^k] < \infty$ for all $k \geq 0$, to conclude that all three terms have finite expectations, and hence

$$E[\tilde{X}^2_{N(t)+1}] = E[\tilde{X}^2_{N(t)}] + 2E[\tilde{X}_{N(t)}(\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)})] + E[(\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)})^2]. \quad (30)$$

Thus, we are done if

$$E[\tilde{X}_{N(t)}(\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)})] = 0.$$

Note that

$$\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)} = \frac{(X_1 - \bar{X}_{N(t)})I(N(t) \geq 1)}{N(t) + 1}$$

and thus

$$E[\tilde{X}_{N(t)}(\tilde{X}_{N(t)+1} - \tilde{X}_{N(t)})] = E\left[\frac{\bar{X}_{N(t)}X_1}{N(t) + 1}\right] - E\left[\frac{\bar{X}^2_{N(t)}}{N(t) + 1}\right]. \quad (31)$$

(Wald's second moment equality again asserts that both terms on the right-hand side are finite.) We now compute the conditional expectation of $\bar{X}^2_{N(t)}/(N(t) + 1)$

$$\begin{aligned} E\left[\frac{\bar{X}^2_{N(t)}}{N(t) + 1}\middle| N(t)\right] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E[X_1X_j \mid N(t)] \frac{I(N(t) \geq \max(i, j))}{(N(t) + 1)N(t)^2} \\ &= E[X_1^2 \mid N(t)] \frac{I(N(t) \geq 1)}{(N(t) + 1)N(t)} \\ &\quad + E[X_1X_2 \mid N(t)] \frac{I(N(t) \geq 2)(N(t) - 1)}{(N(t) + 1)N(t)} \\ &= E\left[X_1 \sum_{i=1}^{N(t)} X_i \frac{I(N(t) \geq 1)}{(N(t) + 1)N(t)}\middle| N(t)\right] \\ &= E\left[\frac{X_1\bar{X}_{N(t)}}{N(t) + 1}\middle| N(t)\right]. \end{aligned} \quad (32)$$

To ensure that all relevant conditional expectations make sense, we can first assume that $X_k \geq 0$ a.s. and then deal with the general case in the usual fashion. By taking expectations of Equation 32, we find that

$$E[\bar{X}_{N(t)}(X_1 - \bar{X}_{N(t)})/(N(t) + 1)] = 0$$

which, by (31), completes the proof.

On the basis of Theorem 4, we therefore recommend the estimator $\tilde{X}_{N(t)}$ over $\tilde{X}_{N(t)+1}$.

Finally, consider the stopping time associated with implementing the estimator $\bar{X}_{N(t)+1}$. The simulation stops at time $T_{N(t)+1} = \tau_1 + \ldots + \tau_{N(t)+1}$ which requires completing the replication in progress at time t. The difference $T_{N(t)+1} - t$ is the excess life in a renewal process which, under appropriate technical conditions, converges in distribution to a finite random variable with mean $E[\tau_1^2]/(2E[\tau_1])$ (see, e.g., p. 195 of Karlin and Taylor). Thus, whereas $E[T_t - t] = o(t^{-p})$ (see Corollary 8), $\lim_{t\to\infty} E[T_{N(t)+1} - t] > 0$. Indeed, in the limit, the expected length of the cycle in progress at time t is always greater than or equal to the expected length of a typical cycle; this, of course, is just a statement of the so-called "inspection paradox." For example, Meketon and Heidelberger showed that in the M/M/1 queue with $\rho = 0.9$, $E[\tau_1] = 10$ but that $\lim_{t\to\infty} E[\tau_{N(t)+1}] = 181$ (here τ_k is the number of customers served during a regenerative cycle). Thus, not only are $\bar{X}_{N(t)}$ and $\tilde{X}_{N(t)}$ preferable to $\bar{X}_{N(t)+1}$ from a statistical viewpoint, they are computationally more efficient as well.

4. GENERAL FUNCTIONS OF MEANS

In this section, we consider the problem of estimating a general function of means, $g(\boldsymbol{\mu})$. For completeness, we begin by stating, without detailed proof, a strong law and central limit theorem for $g(\overline{\mathbf{X}}_{N(t)})$; this theorem is undoubtedly well known and its method of proof is straightforward. A similar central limit theorem for the case of a constant number of observations has, for example, been considered in Section 28.4 of Cramér, and the discussion in Miller points to other references for this technique as well.

Theorem 5.

1. *If* $\mathrm{E}[|X_k(i)|] < \infty$ *for each i and g is continuous in a neighborhood of* $\boldsymbol{\mu}$, *then*

$$\lim_{t\to\infty} g(\overline{\mathbf{X}}_{N(t)}) = g(\boldsymbol{\mu}) \quad a.s.$$

2. *If* $\mathrm{E}[X_k(i)^2] < \infty$ *for each i,* $\mathrm{E}[\tau_k] < \infty$, *g is continuously differentiable in a neighborhood of* $\boldsymbol{\mu}$ *and*

$$\sigma^2 = \mathrm{E}[\tau_k] \sum_{i=1}^{d} \sum_{j=1}^{d} g_i g_j C_{ij}$$

then

$$\sqrt{t}(g(\overline{\mathbf{X}}_{N(t)}) - g(\boldsymbol{\mu})) \Rightarrow N(0, \sigma^2).$$

The proof of the central limit theorem relies on two observations. First, if t is large enough, then $\overline{\mathbf{X}}_{N(t)}$ is close enough to $\boldsymbol{\mu}$ so that

$$g(\overline{\mathbf{X}}_{N(t)}) - g(\boldsymbol{\mu}) = \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\boldsymbol{\xi}(t))(\overline{X}_{N(t)}(i) - \mu_i) \tag{33}$$

where $\boldsymbol{\xi}(t)$ lies on the line segment joining $\boldsymbol{\mu}$ and $\overline{\mathbf{X}}_{N(t)}$ and $(\partial g/\partial x_i)(\boldsymbol{\xi}(t)) \Rightarrow g_i$ for each i. Furthermore, a straightforward application of the Cramér–Wold device to Theorem 17.1 of Billingsley yields the multidimensional central limit theorem

$$\sqrt{t}(\overline{\mathbf{X}}_{N(t)} - \boldsymbol{\mu}) \Rightarrow \mathbf{N}(\mathbf{0}, \mathrm{E}[\tau_k]\mathbf{C}).$$

Standard weak convergence arguments complete the proof.

Note that the above theorem is also valid for both $g(\tilde{\mathbf{X}}_{N(t)})$ and $g(\overline{\mathbf{X}}_{N(t)+1})$ under the identical set of hypotheses. We next turn to the bias expectation of $g(\overline{\mathbf{X}}_{N(t)})$. We begin by establishing uniform integrability and moment convergence of $|\sqrt{t}(\overline{X}_{N(t)}(i) - \mu_i)|^p$. In what follows, we are not looking for minimal moment conditions, but rather simplicity of arguments.

Theorem 6. *If* $\mathrm{E}[|X_k(i)|^{2p+1+\delta}] < \infty$ *and* $\mathrm{E}[\tau_k^{2p+\delta}] < \infty$ *for some* $p \geq 0$ *and* $\delta > 0$, *then*

$$\lim_{t\to\infty} \mathrm{E}[|\sqrt{t}(\overline{X}_{N(t)}(i) - \mu_i)|^p] = \mathrm{E}[|N(0, \mathrm{E}[\tau_k]C_{ii})|^p].$$

Proof. As in Section 2, we can ignore the contribution on the set $\{N(t) = 0\}$. Let

$$\hat{S}_{N(t)}(i) = \sum_{k=1}^{N(t)} (X_k(i) - \mu_i).$$

By Cauchy–Schwarz

$$\mathrm{E}[I(N(t) > 0)|\sqrt{t}(\overline{X}_{N(t)}(i) - \mu_i)|^{p+\epsilon}] = \mathrm{E}\left[\left|\frac{I(N(t) > 0)t}{N(t)} \frac{\hat{S}_{N(t)}(i)}{\sqrt{t}}\right|^{p+\epsilon}\right] \tag{34}$$

$$\leq \mathrm{E}\left[\left|\frac{I(N(t)>0)t}{N(t)}\right|^{2p+2\epsilon}\right]^{1/2} \mathrm{E}\left[\left|\frac{\hat{S}_{N(t)}(i)}{\sqrt{t}}\right|^{2p+2\epsilon}\right]^{1/2}.$$

By Corollary 11, the first term on the right-hand side of Inequality 34 is bounded provided that $\mathrm{E}[\tau_1^{2p+\delta}] < \infty$ and $2\epsilon < \delta$. By Theorem 2 of Chow, Hsiung and Lai, the second term on the right-hand side of Inequality 34 is bounded provided that $\mathrm{E}[\tau_1^{2p+\delta}] < \infty$, $\mathrm{E}[|X_1(i)|^{2p+1+\delta}] < \infty$ and $2\epsilon < \delta$.

Theorem 6 is also valid for $\overline{X}_{N(t)+1}(i)$ and $\tilde{X}_{N(t)}(i)$ under the same hypotheses (actually, by Theorem 2 of Chow, Hsiung and Lai; for $\overline{X}_{N(t)+1}(i)$ we only need $\mathrm{E}[|X_k(i)|^{2p+\delta}] < \infty$ since $N(t) + 1$ is a stopping time).

Theorem 7. *If* $\mathrm{E}[|X_k(i)|^{5+\delta}] < \infty$ *for each i,* $\mathrm{E}[\tau_k^{4+\delta}] < \infty$ *for some* $\delta > 0$, *g is bounded with probability one and twice continuously differentiable in a neighborhood of* $\boldsymbol{\mu}$, *then*

$$\mathrm{E}[g(\overline{\mathbf{X}}_{N(t)})] = g(\boldsymbol{\mu}) + \frac{\mathrm{E}[\tau_k]}{2t} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij} C_{ij} + o\left(\frac{1}{t}\right).$$

Proof. Since g is twice continuously differentiable, if

$$\|\mathbf{x} - \boldsymbol{\mu}\| = \max_{1\leq i\leq d} |x_i - \mu_i| \leq \epsilon$$

then

$$\left|\frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{x})\right| \leq B_\epsilon \quad \text{for some } B_\epsilon < \infty.$$

Let $\Gamma(t)$ be the indicator of the event $\{\| \bar{\mathbf{X}}_{N(t)} - \boldsymbol{\mu} \| \leq \epsilon\}$. Then

$$\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})] = \mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})\Gamma(t)] + \mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})(1 - \Gamma(t))].$$

If $\Gamma(t) = 0$, then $|g(\bar{\mathbf{X}}_{N(t)}) - g(\boldsymbol{\mu})| \leq 2A$ (the constant A is a global bound on g). Thus

$$\begin{aligned} &|\mathrm{E}[(g(\bar{\mathbf{X}}_{N(t)}) - g(\boldsymbol{\mu}))(1 - \Gamma(t))]| \\ &\leq 2A \sum_{i=1}^{d} \mathbf{P}\{|\bar{X}_{N(t)}(i) - \mu_i| > \epsilon\} \\ &= 2A \sum_{i=1}^{d} \mathbf{P}\{\sqrt{t}\,|\bar{X}_{N(t)}(i) - \mu_i| > \sqrt{t}\epsilon\} \\ &\leq \frac{2A}{(\sqrt{t}\epsilon)^{2+\beta}} \sum_{i=1}^{d} \mathrm{E}[(\sqrt{t}\,|\bar{X}_{N(t)}(i) - \mu_i|)^{2+\beta}] \\ &= \mathrm{o}\left(\frac{1}{t}\right) \end{aligned} \tag{35}$$

by Theorem 6 provided that $2\beta < \delta$. Thus, we only have to worry about the contribution of the term when $\Gamma(t) = 1$. However, on this set we have the Taylor series expansion

$$\begin{aligned} g(\bar{\mathbf{X}}_{N(t)}) &= g(\boldsymbol{\mu}) + \sum_{i=1}^{d} g_i(\bar{X}_{N(t)}(i) - \mu_i) \\ &+ \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij}(\boldsymbol{\xi}(t)) \\ &\cdot (\bar{X}_{N(t)}(i) - \mu_i)(\bar{X}_{N(t)}(j) - \mu_j) \end{aligned} \tag{36}$$

where

$$G_{ij}(\boldsymbol{\xi}(t)) = \frac{\partial^2}{\partial x_i \partial x_j} g(\boldsymbol{\xi}(t))$$

and $\boldsymbol{\xi}(t)$ is on the line segment joining $\bar{\mathbf{X}}_{N(t)}$ and $\boldsymbol{\mu}$. From Corollary 5

$$\begin{aligned} &\mathrm{E}[|\bar{X}_{N(t)}(i) - \mu_i|\Gamma(t)] \\ &\leq \mathrm{E}[|\bar{X}_{N(t)}(i) - \mu_i|] = \mathrm{o}(1/t). \end{aligned}$$

Furthermore, since

$$\begin{aligned} &\Gamma(t)G_{ij}(\boldsymbol{\xi}(t))t(\bar{X}_{N(t)}(i) - \mu_i)(\bar{X}_{N(t)}(j) - \mu_j) \\ &\Rightarrow G_{ij}N_iN_j \end{aligned}$$

where N_i and N_j are jointly normal (with means zero and covariance $\mathrm{E}[N_iN_j] = \mathrm{E}[\tau_k]C_{ij}$) and $|G_{ij}(\boldsymbol{\xi}(t))| \leq B_\epsilon$ when $\Gamma(t) = 1$, we have only to establish the uniform integrability of $\Gamma(t)t(\bar{X}_{N(t)}(i) - \mu_i)(\bar{X}_{N(t)}(j) - \mu_j)$. By Cauchy-Schwarz, it suffices to show that

$$\sup_t \mathrm{E}[\Gamma(t)|\sqrt{t}(\bar{X}_{N(t)}(i) - \mu_i)|^{2+\beta}] < \infty \quad \text{for some } \beta > 0 \text{ and each } i.$$

But this follows directly from Theorem 6.

Combining the analysis techniques above with Corollary 7 and Theorem 3 yields the following bias expansions for $g(\tilde{\mathbf{X}}_{N(t)})$ and $g(\bar{\mathbf{X}}_{N(t)+1})$, which are stated without proofs.

Theorem 8. *If* $\mathrm{E}[|X_k(i)|^{5+\delta}] < \infty$ *for each* i, $\mathrm{E}[\tau_k^{4+\delta}] < \infty$ *for some* $\delta > 0$, g *is bounded with probability one and twice continuously differentiable in a neighborhood of* $\boldsymbol{\mu}$, *then*

$$\mathrm{E}[g(\tilde{\mathbf{X}}_{N(t)})] = g(\boldsymbol{\mu}) + \frac{\mathrm{E}[\tau_k]}{2t} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij}C_{ij} + \mathrm{o}\left(\frac{1}{t}\right).$$

Theorem 9. *If* $\mathrm{E}[|X_k(i)|^{4+\delta}] < \infty$ *for each* i, $\mathrm{E}[\tau_k^{4+\delta}] < \infty$ *for some* $\delta > 0$, g *is bounded with probability one and twice continuously differentiable in a neighborhood of* $\boldsymbol{\mu}$, *then*

$$\begin{aligned} \mathrm{E}[g(\bar{\mathbf{X}}_{N(t)+1})] &= g(\boldsymbol{\mu}) + \frac{1}{t} \sum_{i=1}^{d} g_i c_i \\ &+ \frac{\mathrm{E}[\tau_k]}{2t} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij}C_{ij} + \mathrm{o}\left(\frac{1}{t}\right). \end{aligned}$$

The condition that g be bounded can be replaced by other regularity conditions, for example, bounded second derivatives as seen in Equation 36. In the case of ratio estimation, the boundedness conditions can be removed entirely by assuming that $\mathrm{E}[|X_k(i)|^4] < \infty$, $\mathrm{E}[\tau_k^6] < \infty$ and $\mathbf{P}\{\tau_k \geq \epsilon\} = 1$ for some $\epsilon > 0$ (see Meketon and Heidelberger). In the case of ratios, the condition that g be bounded will be satisfied if

$$X_k(i) = \int_{\{\text{Cycle } k\}} f_i(Z_s)\, ds$$

for a bounded function f_i, as often arises in the analysis of a regenerative process $\mathbf{Z} = \{Z_s, s \geq 0\}$.

The bias expansions for $g(\tilde{\mathbf{X}}_{N(t)})$ and $g(\bar{\mathbf{X}}_{N(t)})$ suggest that $1/t$ term can be removed by estimating $\mathrm{E}[\tau_k]$, G_{ij} and C_{ij}. If we expand the dimensionality of $\boldsymbol{\mu}$ to include $\mathrm{E}[\tau_k]$, $\mathrm{E}[X_k(i)X_k(j)]$ and $\mathrm{E}[\tau_kX_k(i)]$, then estimation of

$$h(\boldsymbol{\mu}) \equiv \left(\frac{\mathrm{E}[\tau_k]}{2}\right) \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij} C_{ij}$$

is a special case of estimation of a nonlinear function of means. In particular, under suitable regularity conditions, e.g., if the assumptions of, say, Theorem 8 are satisfied for this expanded $h(\boldsymbol{\mu})$, then $\mathrm{E}[h(\tilde{\mathbf{X}}_{N(t)})] = h(\boldsymbol{\mu}) + \mathrm{o}(1/t)$. Thus, $\mathrm{E}[h(\tilde{\mathbf{X}}_{N(t)})]/t = h(\boldsymbol{\mu})/t + \mathrm{o}(1/t)$ and therefore

$$\mathrm{E}[g(\tilde{\mathbf{X}}_{N(t)})] - \frac{1}{t}\mathrm{E}[h(\tilde{\mathbf{X}}_{N(t)})] = g(\boldsymbol{\mu}) + \mathrm{o}\left(\frac{1}{t}\right). \qquad (37)$$

This Taylor series adjusted estimator is entirely analogous to the Tin estimator (Tin 1965). While it is not necessarily reasonable to assume that the function h satisfies the boundedness condition of Theorem 8, general regularity conditions under which the bias expansion of Equation 37 is valid will not be pursued here.

While this Taylor series adjusted estimator potentially reduces the bias from $\mathrm{o}(1/t)$ to $\mathrm{o}(1/t)$, its performance in practical applications is unclear, even in the case of a constant number of observations (see, e.g., Miller and Iglehart). Furthermore, note that since

$$h(\tilde{\mathbf{X}}_{N(t)})/\sqrt{t} \Rightarrow 0$$

by Theorem 4.1 of Billingsley, the Taylor series adjusted estimator $g(\tilde{\mathbf{X}}_{N(t)}) - h(\tilde{\mathbf{X}}_{N(t)})/t$ obeys the same central limit theorem (Theorem 5) as $g(\tilde{\mathbf{X}}_{N(t)})$.

5. RATIO ESTIMATION

In this section, we apply the results of Section 4 to the case of ratio estimation in regenerative simulations to verify and reinterpret the bias expansions in Meketon and Heidelberger, which were derived using somewhat different methods. We assume that we are interested in estimating a ratio $g(\boldsymbol{\mu}) = \mathrm{E}[X_k(1)]/\mathrm{E}[\tau_k](X_k(2) = \tau_k)$. As mentioned before, this corresponds to the case where both t and τ_k are in units of simulated time. Using the notation of the previous sections, differentiation of g yields

$$g_1 = \frac{1}{\mathrm{E}[\tau_k]}, \quad g_2 = -\frac{\mathrm{E}[X_k(1)]}{\mathrm{E}[\tau_k]^2}$$

$$G_{11} = 0, \quad G_{22} = \frac{2\mathrm{E}[X_k(1)]}{\mathrm{E}[\tau_k]^3}$$

and

$$G_{12} = G_{21} = -\frac{1}{\mathrm{E}[\tau_k]^2}.$$

Furthermore, $C_{11} = \mathrm{Var}[X_k(1)]$, $C_{22} = \mathrm{Var}[\tau_k]$ and $C_{12} = C_{21} = \mathrm{Cov}[X_k(1), \tau_k]$. Thus, from Theorem 7, the bias expansion of $g(\bar{\mathbf{X}}_{N(t)})$ is

$$\begin{aligned}\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)})] &= \frac{\mathrm{E}[X_k(1)]}{\mathrm{E}[\tau_k]} + \frac{\mathrm{E}[\tau_k]}{2t} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij} C_{ij} + \mathrm{o}\left(\frac{1}{t}\right) \\ &= \frac{\mathrm{E}[X_k(1)]}{\mathrm{E}[\tau_k]} + \frac{1}{t}\left(\frac{\mathrm{E}[X_k(1)]\mathrm{Var}[\tau_k]}{\mathrm{E}[\tau_k]^2} - \frac{\mathrm{Cov}[X_k(1), \tau_k]}{\mathrm{E}[\tau_k]}\right) + \mathrm{o}\left(\frac{1}{t}\right) \qquad (38)\end{aligned}$$

which agrees with Meketon and Heidelberger. In addition, $c_1 = \mathrm{Cov}[X_k(1), \tau_k]$ and $c_2 = \mathrm{Var}[\tau_k]$ so that

$$\sum_{i=1}^{d} g_i c_i = \frac{\mathrm{Cov}[X_k(1), \tau_k]}{\mathrm{E}[\tau_k]} - \frac{\mathrm{E}[X_k(1)]\mathrm{Var}[\tau_k]}{\mathrm{E}[\tau_k]^2} \qquad (39)$$

which exactly cancels the $1/t$ term involving second derivatives when computing the bias expansion of $g(\bar{\mathbf{X}}_{N(t)+1})$, as given by Theorem 9, i.e., in this particular case

$$\sum_{i=1}^{d} g_i c_i = -\frac{\mathrm{E}[\tau_k]}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} G_{ij} C_{ij}. \qquad (40)$$

Thus, $\mathrm{E}[g(\bar{\mathbf{X}}_{N(t)+1})] = \mathrm{E}[X_k(1)]/\mathrm{E}[\tau_k] + \mathrm{o}(1/t)$ which also agrees with Meketon and Heidelberger.

As seen from these Taylor series expansions, the bias reduction, when using $N(t) + 1$ cycles in the ratio estimator, depends critically upon:

1. Since $X_2(k) = \tau_k$, the c_i's and the C_{ij}'s are related. More specifically, $c_1 = C_{12} = C_{21} = \mathrm{Cov}[X_k(1), \tau_k]$ and $c_2 = C_{22} = \mathrm{Var}[\tau_k]$.
2. The very particular form of the ratio function g as represented by the relationships $G_{11} = 0$, $g_1 = -(\mathrm{E}[\tau_k]/2)(G_{12} + G_{21})$ and $g_2 = -(\mathrm{E}[\tau_k]/2)G_{22}$.

Without these special relationships, cancellation of the first and second derivative terms of order $1/t$ in the bias expansion of $g(\bar{\mathbf{X}}_{N(t)+1})$ as expressed in Equation 40 would not, in general, occur.

6. SUMMARY

This paper has examined the bias characteristics of estimates that typically arise in Monte Carlo simulations. If there is a budget constraint t that represents the maximum amount of computer time available for the simulation, then the number of observations (replications) $N(t)$, completed by time t is a random variable. This randomness can introduce bias, even in the case of estimating a simple sample mean. However, this bias was, surprisingly, shown to be of at most order $o(1/t)$ rather than of order $1/t$ as we had originally thought, i.e., $E[\bar{X}_{N(t)}] = E[X_1] + o(1/t)$. Under suitable moment conditions, $E[\bar{X}_{N(t)}]$ approaches $E[X_1]$ at a much faster rate than $1/t$. Furthermore, if the replication lengths are bounded and if t is greater than the bound, then $\bar{X}_{N(t)}$ is unbiased. In addition, the exact expression for $E[\bar{X}_{N(t)}]$ suggests a simple adjustment to $\bar{X}_{N(t)}$ that results in an unbiased estimator. This adjusted estimator, $\tilde{X}_{N(t)}$, is the sample mean of $\max(1, N(t))$ replications and requires completing the replication in progress at time t only if no replications have yet been completed.

On the other hand, the bias of the sample mean using $N(t) + 1$ observations, which is obtained by always completing simulation of the replication in progress at time t, was shown to be of order $1/t$, i.e., $E[\bar{X}_{N(t)+1}] = E[X_1] + c/t + o(1/t)$ where the constant c was explicitly identified. In contrast to $E[\bar{X}_{N(t)}]$, $E[\bar{X}_{N(t)+1}]$ always approaches $E[X_1]$ at rate $1/t$, no matter how many additional moments are assumed finite.

Bias expressions for estimating a general function of means, $g(\boldsymbol{\mu})$, were then obtained for both $N(t)$ and $N(t) + 1$ replications. The leading term in these bias expansions, which are based on Taylor series expansions, are in general of order $1/t$ where the coefficients of $1/t$ are explicitly identified. To order $1/t$, the bias in $g(\bar{\mathbf{X}}_{N(t)})$ (or $g(\tilde{\mathbf{X}}_{N(t)})$) is entirely due to the nonlinearity of the function g. The bias expansion of $g(\bar{\mathbf{X}}_{N(t)+1})$ contains an extra term representing the $1/t$ bias in the sample means $\bar{\mathbf{X}}_{N(t)+1}$ due to the random number of observations. A Tin-style adjustment to the estimator $g(\tilde{\mathbf{X}}_{N(t)})$ based on the Taylor series expansion was proposed to reduce the bias.

These results were then applied to the case of ratio estimation in regenerative simulation. For this case, a previously proposed bias reducing technique of using the ratio estimator with $N(t) + 1$ cycles eliminates the bias of order $1/t$. Using the Taylor series bias expansions, it was shown that this bias reduction comes about because of very special circumstances and cannot be expected in more general situations.

The results of this paper form the bias for estimation procedures when independent replications are run in parallel on multiple processor computers. A partial treatment may be found in Heidelberger, and a more complete treatment is contained in Glynn and Heidelberger (1990).

ACKNOWLEDGMENT

This work was partially performed while Peter Glynn was a visiting scientist at the IBM T. J. Watson Research Center. The work of Peter Glynn was also supported by the U.S. Army under contract DAAG29-84-K-0030.

REFERENCES

BEALE, E. M. L. 1962. Some Uses of Computers in Operational Research. *Indust. Organ.* **31,** 27–28.

BILLINGSLEY, P. 1968. *Convergence of Probability Measures.* John Wiley, New York.

BROWN, M., AND H. SOLOMON. 1975. Second Order Approximations for the Variance of a Renewal Reward Process. *Stoch. Proc. Appl.* **3,** 301–314.

CHOW, Y. S., C. A. HSIUNG AND T. L. LAI. 1979. Extended Renewal Theory and Moment Convergence in Anscombe's Theorem. *Ann. Prob.* **7,** 304–318.

CHOW, Y. S., H. ROBBINS AND D. SIEGMUND. 1971. *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston.

CHUNG, K. L. 1974. *A Course in Probability Theory,* 2nd ed. Academic Press, New York.

COX, D. R. 1962. *Renewal Theory.* Methuen and Co., London.

CRAMÉR, H. 1946. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, N.J.

CRANE, M. A., AND D. L. IGLEHART. 1975. Simulating Stable Stochastic Systems, III: Regenerative Processes and Discrete Event Simulations. *Opns. Res.* **23,** 33–45.

GLYNN, P. W. 1987. A Low Bias Steady-State Estimator for Equilibrium Processes. Technical Report, Department of Industrial Engineering, University of Wisconsin, Madison.

GLYNN, P. W., AND P. HEIDELBERGER. 1990. Analysis of Parallel, Replicated Simulations Under a Completion Time Constraint. IBM Research Report RC15466, Yorktown Heights, N.Y. (submitted for publication).

HEIDELBERGER, P. 1988. Discrete Event Simulations and Parallel Processing: Statistical Properties. *SIAM J. Sci. Statist. Comput.* **9,** 1114–1132.

IGLEHART, D. L. 1975. Simulating Stable Stochastic Systems, V: Comparison of Ratio Estimators. *Naval Res. Logis. Quart.* **22,** 553–565.

KARLIN, S., AND H. M. TAYLOR. 1975. *A First Course in Stochastic Processes,* 2nd ed. Academic Press, New York.

KREMERS, W. K. 1986. Completeness and Unbiased

Estimation for Sum-Quota Sampling. *J. Am. Statist. Assoc.* **81,** 1070–1073.

LAVENBERG, S. S., AND C. H. SAUER. 1975. Sequential Stopping Rules for the Regenerative Method of Simulation. *IBM J. Res. Develop.* **21,** 545–558.

LEHMANN, E. L. 1983. *Theory of Point Estimation.* John Wiley, New York.

MEKETON, M. S., AND P. HEIDELBERGER. 1982. A Renewal Theoretic Approach to Bias Reduction in Regenerative Simulations. *Mgmt. Sci.* **28,** 173–181.

MILLER, R. G. 1974. The Jackknife—A Review. *Biometrika* **61,** 1–15.

MURTHY, V. K. 1974. *The General Point Process, Applications to Structural Fatigue, Bioscience, and Medical Research.* Addison-Wesley, Reading, Mass.

PATHAK, P. K. 1976. Unbiased Estimation in Fixed Cost Sequential Sampling Schemes. *Ann. Statist.* **4,** 1012–1017.

ROSS, S. M. 1983. *Stochastic Processes.* John Wiley, New York.

SMITH, W. L. 1955. Regenerative Stochastic Processes. *Proc. Roy. Soc. A.* **232,** 6–31.

SMITH, W. L. 1958. Renewal Theory and Its Ramifications. *J. Roy. Statist. Soc. B* **20,** 243–302.

SMITH, W. L. 1959. Cumulants of Renewal Processes. *Biometrika* **46,** 1–29.

TIN, M. 1965. Comparison of Some Ratio Estimators. *J. Am. Statist. Assoc.* **60,** 294–307.