ESTIMATING TIME AVERAGES VIA RANDOMLY-SPACED OBSERVATIONS

Bennett L. Fox and Peter W. Glynn

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

January 1984

(Received October 14, 1983)

Approved for public release
Distribution unlimited

Sponsored by
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709
ESTIMATING TIME AVERAGES VIA RANDOMLY-SPACED OBSERVATIONS

Bennett L. Fox* and Peter W. Glynn*

Technical Summary Report #2638
January 1984

ABSTRACT

To estimate continuous-time averages via randomly-spaced observations of discrete-event systems, we develop a point-process framework and use it to generalize both regenerative and stationary-process oriented simulation methodologies. We give consistent estimators, central limit theorems, and an effective bias-reducing jackknife. The impact on indirect estimation of transaction (customer) averages is discussed.

AMS (MOS) Subject Classifications: 62M15, 65C05, 68J05, 60G55

Key Words: Simulation, point processes

Work Unit Number 5 (Optimization and Large Scale Systems)

*Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. Research also supported by

1The Natural Sciences and Engineering Research Council of Canada.
2The Graduate School of the University of Wisconsin-Madison.
SIGNIFICANCE AND EXPLANATION

In many stochastic systems, one is interested in estimating steady-state expected values. When Monte Carlo simulation is used to estimate such parameters, an assessment of accuracy, in the form of confidence intervals, is often required. Most procedures for producing such confidence intervals require that the simulation be sampled so that the time increments between observations are all equal. This is difficult to accomplish in a discrete-event simulation, since the clock which drives the simulation is incremented in a random fashion. Our purpose, in this paper, is to show how methods for dealing with equally spaced observations can be adapted to run on the random time scale of the driving clock for the simulation.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.
ESTIMATING TIME AVERAGES VIA RANDOMLY-SPACED OBSERVATIONS

Bennett L. Fox* and Peter W. Glynn*

1. INTRODUCTION.

Let \( 0 = T_0 < T_1 < \ldots \) be event times. Though the associated sequence \( \{X_1, X_2, \ldots \} \) with \( X_k \in \mathbb{R}^d \) is not necessarily an embedded Markov chain, we call \( X_k \) the state at time \( T_k \) - somewhat abusing the term. To define the state \( X(t) \) at an arbitrary time \( t \), interpolate:

\[
X(t) = \sum_{k=0}^{n} X_k I_{(T_k, T_{k+1}]}(t)
\]

where the indicator \( I_{(a,b]} \) is 1 or 0 depending on whether or not \( t \in (a,b] \). For this definition to make sense, every state change must correspond to an event time. The state does not change continuously. It jumps at discrete (possibly random) times. In other words, we have a discrete-event system. Let \( f \) be a real-valued function. Put

\[
r(t) = \int_0^t f(X(s)) ds
\]

We solve the steady-state problem: estimate the limit (when it exists)

\[
r = \lim_{t \to \infty} r(t)
\]

and construct confidence intervals for \( r \).

To do this, we develop a point-process framework and use it to generalize both regenerative and stationary-process oriented simulation methodologies. Simply averaging the \( X_k \)'s generally inconsistently estimates \( r \). The \( T_k \)'s are not necessarily regeneration times. We work with generally dependent observations, in contrast to regenerative approaches.

*Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
Research also supported by
1The Natural Sciences and Engineering Research Council of Canada
2The Graduate School of the University of Wisconsin-Madison.
1.1 Shifts. We use the set of functions \( \omega : [0,\infty) \times \mathbb{R}^d \) that are right continuous and have left limits to describe the sample space \( \Omega \). Define \( X \) via

\[
X(x) = X(x, \omega) = \omega(x)
\]

for \( \omega \in \Omega \). For any random variable \( R : \Omega \times [0,\infty) \rightarrow \mathbb{R}^d \), define a right shift via

\[
\theta_R = \theta_R(\omega) = \omega(R(\omega) + \cdot)
\]

Let \( A \subset \mathbb{R}^d \), put \( S_0 = 0 \), and set

\[
S_{k+1} = \inf\{t > S_k : X(t-) \neq X(t), X(t) \in A\}
\]

where \( X(t-) \) is the left limit of \( X \) at \( t \). If \( A = \mathbb{R}^d \), then \( S_1 = T_A \). Define

\[
P_n(x \in \cdot) = P(X \in \theta_{S_n} \in \cdot)
\]

and

\[
P_{\infty}(x \in \cdot) = P(X \in \theta_{\infty} \in \cdot)
\]

For most simulations where the steady-state limit \( r \) exists, limit probabilities \( \hat{P} \) and \( \bar{P} \) exist such that weak convergence holds:

\[
P_n \Rightarrow \hat{P}
\]

and

\[
P_{\infty} \Rightarrow \bar{P}
\]

where

\[
\hat{P}(X \in \theta_{S_1} \in \cdot) = \hat{P}(X \in \cdot)
\]

and (shift invariance)

\[
\bar{P}(X \in \theta_{t} \in \cdot) = \bar{P}(X \in \cdot)
\]

for \( t > 0 \).

**Example 1.1.** Let \( X \) be a delayed (resp., non-delayed) regenerative process under \( P \) (resp., \( \hat{P} \)). Assume that \( X \) regenerates when it hits \( A \). Then \( P_n = \hat{P} \) for \( n > 1 \) though generally \( P_0 \neq \hat{P} \).

**Example 1.2.** Let \( X \) be as in example 1.1. If the regenerative process is positive recurrent and the regeneration-spacing distribution has a non-trivial Lebesgue component, then (1.10) and (1.12) hold (see Miller (1972)). Also, \( \bar{P} \) and \( \hat{P} \) are related by
1.13) \[ \mathbb{P}(X \in \cdot) = \frac{1}{\mathbb{E}_1} \int_{\mathbb{E}_1} \mathbb{P}(X \in \cdot; S_1 > t)dt \]

where \( \mathbb{E} \) denotes expectation under \( \mathbb{P} \). In Section 3 we show that (1.13) holds more generally.

A process \( X \) satisfying

\[ \mathbb{P}(X \in \cdot; S_1) = \mathbb{P}(X \in \cdot) \]

is synchronous with respect to the imbedded point process sequence \( \{S_n\} \).

**Example 1.3.** Suppose that \( \{X_n\} \) is a stationary sequence with \( X_n = T_n = n \). Then \( X(t) \) is synchronous with respect to \( \{S_n\} \).

From examples 1.1 and 1.3 we see that synchronous processes generalize both non-delayed regenerative processes and stationary sequences. Assume that the simulator (somehow) chooses the time origin so that everything representing the initial "transient" phase is to its left. This is trivial for regenerative processes but not in general. This deletion assumption translates mathematically as: (1.14) holds.

Let

\[ \sigma_n = S_{n+1} - S_n \]

and

\[ X_n(t) = \begin{cases} X(S_n + t); & 0 < t < \sigma_n \\ = & ; t > \sigma_n \end{cases} \]

We assume that

\[ \{X_n\} \text{ is a } \phi \text{-mixing sequence (Billingsley (1968), pp. 166-168) with } \phi \text{-mixing coefficients satisfying} \]

\[ \sum_{k=1}^{\infty} \phi_k < \infty \]

and

\[ \mathbb{E}(X_n(\lfloor t \rfloor)^2 + \alpha_n^2) < \infty \]

where
\[ Y_n(f) = \int_0^{\sigma_n} f(X_n(t))dt \]

\[ S_{n+1} = \int_{S_n}^{S_{n+1}} f(X(t))dt. \]

The cycle sequence \{X_n\} is not necessarily iid. If it were, we could simply apply regenerative methodology. Three more definitions set the stage:

\[ \tilde{Y}_n(f) = \sum_{k=0}^{n-1} Y_k(f)/n \]

\[ \tilde{\alpha}_n = S_n/n \]

\[ r_n = \tilde{Y}_n(f)/\tilde{\alpha}_n. \]

1.2. Preview of results. In Section 2 we assume that (1.14), (1.17), and (1.18) hold, usually without further explicit mention. We show that

\[ r_n + r \]

\[ r(t) + r \]

\[ r = \frac{EY_0(f)}{E\alpha_0} \]

and generalize the "inspection paradox"

\[ \frac{1}{t} \int_0^t I_{[x, \infty)}(\alpha_n(s))ds + \int_x^\infty P(\alpha_0 > s)ds/E\alpha_0. \]

The left side of (1.26) is the proportion of time over [0,t] that the cycle in progress had length at least x. Its limit, the right side of (1.26), is precisely that for the regenerative case; e.g., see Bratley, Fox, and Schrage ([1983], problem 3.7.41 or Heyman and Sobel ([1982], §5.5). Thus, for synchronous processes the cycle in progress at time t tends to be longer than average, thereby biasing \( r(t) \).

Next, Section 2 proves a central limit theorem (CLT);

\[ n^{1/2} (r_n - r) \rightarrow \sigma N(0,1)/\alpha_0 \]

\[ n^{1/2} (r(t) - r) \rightarrow \sigma N(0,1)/(\alpha_0)^{1/2} \]
where $N(0,1)$ is a zero-mean, unit-variance, normal random variable,

\[(1.29) \quad \sigma^2 = \mathbb{E}_0 \hat{\sigma}^2 + 2 \sum_{k=1}^{\infty} \mathbb{E}_0 \hat{\sigma} \mathbb{E}_0 \hat{\sigma}
\]

and

\[(1.30) \quad \hat{f}(x) = f(x) - r.
\]

In other words, $Y_k(f) = Y_k(f) - a_k r$. If the synchronous process $X$ is either regenerative or stationary with $S_n = n$, the CLT simplifies. For the former, the second term in (1.29) corresponding to covariances vanishes. For the latter, $a_n \neq 1$. Also (1.31)

\[(1.31) \quad \text{when } A = \mathbb{R}^d,
\]

\[\sigma^2 = \mathbb{E}_0 [\hat{\tau}_0^2 + 2 \sum_{k=1}^{\infty} \hat{\tau}_0 \hat{\tau}_k \hat{\tau}_k]
\]

where

\[\hat{\tau}_k = \hat{\tau}_{k+1} - \hat{\tau}_k.
\]

To estimate $\sigma^2$ in the regenerative case, one uses the knowledge that the covariance terms in (1.29) vanish, to construct the estimator $\sigma^2_n$,

\[\sigma^2_n = \frac{1}{n-1} \sum_{k=1}^{n} \left( Y_k(f) - \frac{1}{n} y_k \right)^2
\]

the estimator $\sigma^2_n$ is easily shown to be strongly consistent for $\sigma^2$.

In the general case, estimation of $\sigma^2$ is more complicated. The parameter $\sigma^2$ can be expressed in the form

\[(1.32) \quad \sigma^2 = \left\{ \text{var } Y_0(f) + 2 \sum_{k=1}^{\infty} \text{cov}(Y_0(f), Y_k(f)) + \right\}
\]

\[- r \left[ \text{cov}(Y_0(f), a_k) + 2 \sum_{k=1}^{\infty} \text{cov}(Y_0(f), a_k) \right]
\]

\[- r \left[ \text{cov}(a_0, Y_0(f)) + 2 \sum_{k=1}^{\infty} \text{cov}(a_0, Y_k(f)) \right]
\]

\[+ r^2 \left[ \text{var } a_0 + 2 \sum_{k=1}^{\infty} \text{cov}(a_0, a_k) \right]
\]
the four bracketed terms appearing in (1.32) can be consistently estimated by
\[ c_{1n}^{2}, c_{2n}^{2}, c_{3n}^{2}, c_{4n}^{2} \] (say); this can be accomplished by standard techniques (e.g. batch means, spectral methods, autoregressive procedures). Such methods include parameters (e.g. batch size, spectral window, autoregressive order) which must be keyed to sample size. In any case, the estimator \( \sigma_n^2 \) is then given by
\[
\sigma_n^2 = c_{1n}^2 - c_{2n}^2 - c_{3n}^2 + c_{4n}^2.
\]
When using \( r(t) \), replace \( n \) by
\[
N(t) = \# S_1's \text{ observed in } (0, t).
\]
Glynn and Iglehart ([1981], Jow ([1982], pp. 54-56), and Streller ([1980], Theorem 3.2) prove essentially (1.27) but under significantly different hypotheses.

From (1.27) and (1.28) we get the respective confidence intervals
\[
[r_n - z_0 \sigma_n \sqrt{n}, r_n + z_0 \sigma_n \sqrt{n}]
\]
and
\[
[r(t) - z_0 \sigma_n(t)/(t_0 \sigma_n(t))^{1/2}, r(t) + z_0 \sigma_n(t)/(t_0 \sigma_n(t))^{1/2}]
\]
where the percentile \( z_0 \) is the unique solution of \( P(N(0, 1) < z_0) = 1 - \delta/2 \). If \( \sigma_n \) is a strongly consistent estimator for \( \sigma \), then these are asymptotically exact \( 100(1 - \delta)% \) confidence intervals for \( r \).

Under an additional assumption
\[
\text{(1.34) there exists } K > 0 \text{ such that } P(a_0 < K) = 1
\]
and \( \sup \{|f(x)| : x \in R^3\} < K \)

Section 2 proves that
\[
\text{(1.35) } E\tau_n = r - \beta/(Ea_0^2 + O(1/n))
\]
and finds an expression for \( \beta \). The form of (1.35) motivates a jackknife:
\[
\widetilde{r}_{2n} = 2r_{2n} - (r(0, n - 1) + r(n, 2n - 1))/2
\]
where
\[
\text{(1.36) } r(a, b) = \left( \sum_{j=a}^{b} Y_j(f) \right) / \left( \sum_{j=a}^{b} a_j \right)
\]
and
\[
\text{(1.37) } (a, b) = \left( \sum_{j=a}^{b} Y_j(f) \right) / \left( \sum_{j=a}^{b} a_j \right)
\]
reducing bias by an order of magnitude.

Sometimes there is no bias. If \( S_n = T_n = n \), then \( \text{Er}_n = r \).

For another case, let \( A = \mathbb{N}^d \) and suppose that the \( T_k \)'s see time averages. For the latter, suppose that

1. \( \text{N}(t) \) is a stationary Poisson process
2. for each \( t > 0 \), \((X(s) : 0 < s < t)\) is independent of \( \{\text{N}(t+u) - \text{N}(t) : u > 0\} \).

Wolff (1982) shows that, if \( r(t) + r \) a.s., then

\[
\tilde{r}_n = \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) + r \quad \text{a.s.}
\]

Since the summands are identically distributed, \( \tilde{r}_n = r \). Section 2 concludes by proving that

\[
\frac{\sqrt{2n}}{r_2n} (\tilde{r}_2n - r) \Rightarrow \mathcal{N}(0,1)/\mathbb{E}r_0.
\]

Combining (1.27) and (1.40), we see that \( \tilde{r}_2n \) and \( r_2n \) have the same asymptotic variance. So our jackknife reduces bias without increasing variance.

Section 3 begins by proving that

\[
\frac{1}{c} \int_X P[X \in \mathbb{S} \in \cdot] ds + P[X \in \cdot]
\]

where \( \tilde{P} \) satisfies (1.13), still assuming that (1.14), (1.17), and (1.18) hold. The point is that (1.13) remains valid when \( X \) is (merely) synchronous. Franken et al. [(1982), Chapter 1] show that it holds under even weaker conditions.

When \( X \) is regenerative, \( \tilde{P}(X \in \cdot) = \tilde{P}(X \in \mathbb{S}_1 \in \cdot) \), "inverting" (1.13). This simple inversion generally fails, assuming only (1.9)-(1.12). Recall that under \( \tilde{P} \), the cycle trapping a fixed time tends to be longer than a typical cycle. In the regenerative case, only the actual trapping cycle is affected, due to independence of cycles. In general, however, neighboring cycles are also affected due to cycle correlation. Franken
et al. (1982), p. 23] show that the general inverse to (1.13) is

\[ \hat{P}(X \in \cdot) = \sum_{k=1}^{\infty} \hat{P}(X \in \cdot | S_k < 1) \]

where \( \lambda = \mathbb{E}[1] \) and \( \mathbb{E} \) denotes expectation under \( \hat{P} \). Comparing (1.11) and (1.14), we see that \( X \) is synchronous under \( \hat{P} \), the Palm distribution of \( \hat{P} \). Here \( \lambda \) is the intensity of \( \hat{P} \), and \( X \) is stationary under \( \hat{P} \).

Chapter 1 of Franken et al. (1982) thoroughly discusses the relationships between \( \hat{P} \) and \( \hat{P} \) and proves an intuitive alternative to (1.42):

\[ \hat{P}(X \in \cdot) = \lim_{h \to 0} \mathbb{E}[X \in \cdot | S_1 < h] \]

This shows that the Palm distribution \( \hat{P} \) is the stationary distribution \( \hat{P} \) conditioned on hitting \( A \) at time 0.

We want (1.12) to hold with \( \hat{P} \) replacing \( \hat{P} \). This means the simulator (somehow) deletes the entire transient phase, choosing the time origin so that this phase is to its left. Typically this is impossible to do exactly in practice, but to proceed mathematically we assume it has been done exactly. This translates as

\[ \hat{P}(X \in \cdot) = \hat{P}(X \in \cdot) \]

holds. This is stronger than (1.14) because \( \mathbb{E}X(t) = t \) under (1.44) but not generally under (1.14). In fact (1.44) usually holds literally only if the initial state is generated by the (generally unknown) stationary distribution \( \hat{P} \). Even for regenerative processes, where synchronization via (1.14) is trivial, making (1.44) hold even to a first approximation is generally hard. Nevertheless, in the rest of this section and in Section 3, usually without further explicit mention, we assume that (1.44) holds and that, under \( \hat{P} \) but not necessarily under \( \hat{P} \), (1.14), (1.17), and (1.18) hold. By contrast, Section 2 assumes that (1.14), (1.17), and (1.18) hold under \( \hat{P} \).

Several counterparts to results in Section 2 are proved in Section 3. There we show that, under \( \hat{P} \), \( r(t) \) is (still) strongly consistent and

\[ \sqrt{t} (r(t) - r) \to \mathcal{N}(0,1)/(\mathbb{E}r_0)^{1/2} \]
where

\[ \sigma^2 = \sum_{k=1}^{\infty} \mathbb{E} \hat{Y}_k(f)^2 + 2 \sum_{k=1}^{\infty} \mathbb{E} \hat{Y}_k(f) \hat{Y}_k(f) = \mathbb{E} \hat{Y}_k(f)^2 \cdot \]

This CLT corresponds to (1.28) with \( \mathbb{E} \) (expectation under \( \hat{P} \)) replacing \( \mathbb{E} \). Estimate \( \sigma^2 \) and construct confidence intervals just as before.

The stationarity assumptions (1.14) and (1.44) can be significantly relaxed. Our CLT's can be extended to include certain nonstationary processes by appealing to Pillingeley [(1968), Theorem 20.2]. Furthermore, our discussion carries over to the steady-state estimation problem for

\[ r = \lim_{t \to 0} \frac{1}{t} \sum_{k=0}^{\infty} f(X(U_k)) I_{[0,U_k]}(t) \]

(when the limit exists), where the \( U_k \)'s are an increasing sequence of random times. Such limits are of interest, for example, in queues where lump-sum rewards are payed out to the server at customer departure epochs. Our arguments go through provided that one modifies the definition of \( Y_n(f) \) to

\[ Y_n(f) = \sum_{j=0}^{n} f(X(U_j)) I_{[S_n,S_{n+1})}(U_j) . \]

1.4 Transactions. To essentially every time average \( r \) there corresponds a transaction (customer) average \( s \) and conversely. Heyman and Stidham (1980) and Heyman and Sobel [(1982), §11.3] establish this correspondence explicitly. Thus, every estimator of \( r \) yields an indirect estimator of \( s \) and conversely. In our framework, Section 4 indicates that on balance estimating \( s \) indirectly is often better than estimating it directly. This conclusion appears contrary to some folklore.

Readers wishing to skip to Section 4 can do so without loss of continuity.
2. RESULTS AND PROOFS; I.

Throughout this section we assume that (1.14), (1.17), and (1.18) hold under $P$.

THEOREM 1. Formulas (1.23), (1.24), and (1.25) hold a.s.

Proof. First, observe that $(Y_n(f), a_n)$ is trivially a functional of $X_n(*)$ and thus

$((Y_n(f), a_n) : n > 0)$ is $\phi$-mixing with the same mixing coefficients as the $X_n$'s. Since

any $\phi$-mixing sequence is ergodic (Lamperti (1977), pp. 95-96), apply Birkhoff's ergodic

theorem (Heyman and Sobel (1982), p. 366) to conclude that

(2.1) $\lim_{n \to \infty} Y_n(f) = EY_0(f)$ a.s.

(2.2) $\lim_{n \to \infty} a_n = Ea_0$ a.s.

proving (1.23) with $r$ given by (1.25). With $N(t)$ defined by (1.33),

(2.3) $\frac{S_N(t)/N(t)}{t/N(t)} < t/N(t) < \frac{S_{N(t)+1}/N(t)}{t/N(t)}$

from the definitions. From $N(t) \to \infty$ a.s. and (2.2), the extreme terms of (2.3) converge
to $Ea_0$ a.s.; as $t/N(t)$ gets squeezed, it converges to $Ea_0$ a.s. For nonnegative $f$,

(2.4) $\frac{1}{t} \left( \frac{N(t)}{t} \right) Y_n(f) < r(t) < \frac{1}{t} \left( \frac{N(t)+1}{t} \right) Y_{n+1}(f)$.

Apply $N(t) \to \infty$ a.s., (2.1), and $N(t)/t \to Ea_0$ a.s. to quickly see that the extreme
terms of (2.4) converge to $r$ given by (1.25), proving (1.24) for $f > 0$. Split a
general $f$ into its positive and negative parts, apply (1.24) to each, and recombine. To
justify the last step, use $(E[Y_1(f)]^2 < EY_0(f)^2 = \cdots$ by (1.18).

THEOREM 2. Formula (1.26) holds.

Proof. Observe that

(2.5) $\frac{1}{n} \int_0^n I_{[x, \infty]}(a_N(s))ds = \frac{1}{n} \sum_{k=0}^{n} a_k I_{[x, \infty]}(a_k) = E(a_0; a_0 > x)$ a.s.
(\mathbb{E}[X|A] \leq \mathbb{E}X_A$ where $I_A$ is $1$ or $0$ depending on whether or not $A$ occurs) and then mimic the proof of Theorem 1.

**THEOREM 3.** Formulas (1.27) and 1.28 hold.

**Proof.** The sequence $\{Y_k(f)\}$ is \(\phi\)-mixing with zero mean. It satisfies the conditions of Billingsley (1968), Theorem 20.1, and hence

\begin{equation}
\sqrt{n} \frac{Y_n(f)}{\mathbb{E} \mathbb{X}^2} \to N(0,1)
\end{equation}

as $n \to \infty$; for $\sigma = 0$, see Billingsley (1968), p. 177. Rewrite (2.6) as

\begin{equation}
\sqrt{n} \frac{S_n(r_n - r)}{\mathbb{E} \mathbb{X}^2} \to N(0,1).
\end{equation}

Now (1.27) follows from (2.7), (2.2), and a routine application of the converging-together lemma (Billingsley 1968, p. 25).

For (1.28), let \(\{t_n\}\) be an arbitrary sequence converging to $\frac{r}{t}$. Apply Billingsley's (1968), p. 146) random time change theorem to the weak invariance principle version of (2.6) to get

\begin{equation}
\sqrt{W(t_k)} \frac{Y_{N(t_k)}(f)}{\mathbb{E} \mathbb{X}^2} \to N(0,1)
\end{equation}

as $k \to \infty$. Another application of the converging-together lemma then yields

\begin{equation}
\sqrt{t_k} \frac{S_{N(t_k)}(r_{t_k}) - r_{t_k}}{\mathbb{E} \mathbb{X}^2} \to N(0,1)/\mathbb{E} \mathbb{X}^2
\end{equation}

using $W(t)/t = \mathbb{E} \mathbb{X}^2$ proved just after (2.3) and $\sqrt{W(t_k)} = \sqrt{t_k} \mathbb{E} \mathbb{X}^2/\mathbb{E} \mathbb{X}^2$. Clearly

\begin{equation}
\sqrt{t_k} \frac{S_{N(t_k)}(r_{t_k}) - r_{t_k}}{\mathbb{E} \mathbb{X}^2} < \sqrt{t_k} \frac{Y_{N(t_k)}(f)}{\mathbb{E} \mathbb{X}^2}.
\end{equation}

Combining a standard Borel-Cantelli argument and $\mathbb{E}Y_k(|f|)^2 < \infty$ proves that

\begin{equation}
Y_k(|f|)^2 \sqrt{t_k} \to 0 \quad \text{a.s.}
\end{equation}

But the right-hand side of (2.10) can be rewritten as

\begin{equation}
(t_k/\mathbb{E}X) \frac{Y_{N(t_k)}(|f|)/\mathbb{E}X}{(N(t_k)/\mathbb{E}X)(N(t_k)/\mathbb{E}X)}
\end{equation}

which, by (2.11), converges to $(\mathbb{E}X)^{-1} = 0$. Apply the converging-together lemma.
to (2.9) and (2.10) to get (1.24) for \( t \) going to \( \infty \) through the sequence \( t_k \). Because that sequence is arbitrary, (1.24) holds without qualification (see Billingsley (1968), p. 16).

Now also assuming (1.34) we have

**THEOREM 4.** Formula (1.35) holds with

\[
\beta = E_0 Y_0 (\hat{f}) + \sum_{k=1}^{\infty} (E_0 Y_k (\hat{f}) + E_0 (\hat{f}) \alpha_k). \tag{2.12}
\]

**Proof.** The infinite series for \( \beta \) converges absolutely by a remark of Billingsley (1968), p. 177. Turning to the bias expansion itself, observe that

\[
r_n - r = \tilde{f}(\hat{g}) g(\hat{a}_n) \tag{2.13}
\]

where

\[
g(x) = 1/x. \tag{2.14}
\]

Choosing \( \varepsilon \) small enough,

\[
\sup \{g'(x) : |x - E_0| < \varepsilon\} = M < \infty. \tag{2.15}
\]

Letting event \( B_n = \{|\tilde{a}_n - E_0| < \varepsilon\}, \) use (1.34) and then Chebyshev's inequality to get

\[
|E[r_n - r; B_n]| < KE(\tilde{a}_n - E_0)^4 / \varepsilon^4. \tag{2.16}
\]

By Billingsley ([1968], lemma 4, p. 172),

\[
E(\tilde{a}_n - E_0)^4 = O(n^{-2}). \tag{2.17}
\]

Next

\[
E[r_n - r] = E[r_n - r; B_n] + E[r_n - r; \tilde{a}_n]. \tag{2.18}
\]

Combining (2.16)-(2.17) gives

\[
E[r_n - r] = E[r_n - r; B_n] + O(n^{-2}). \tag{2.19}
\]

On the event \( B_n \), expand \( g(\tilde{a}_n) \) in a Taylor series around \( E_0 \) to get

\[
E[r_n - r; B_n] = S + T
\]

where
(2.20) \[ s = \mathbb{E}(\hat{f}_n^2)\left[ \frac{1}{\mathbb{E}r_0^2} \right] (\mathbb{E} \xi_n - \mathbb{E} \xi_0) ; \frac{b}{n} \] 

(2.21) \[ T = \mathbb{E}(\hat{f}_n^2) g^2(\xi_n^2) (\mathbb{E} \xi_n - \mathbb{E} \xi_0)^2 / 2 ; \mathbb{E} \] 

and

(2.22) \[ |\xi_n - \mathbb{E} \xi_0| < \varepsilon . \]

Apply Cauchy-Schwartz and then Billingsley ([1968], Lemmas 3 and 4, p. 172) to get

(2.23) \[ T < \mathbb{E}(\hat{f}_n^2) \xi_n^2 / 2 (\mathbb{E} \xi_n - \mathbb{E} \xi_0)^4 )^{1/2} = O(n^{-3/2}) . \]

An argument similar to that justifying (2.18) gives

(2.24) \[ s = -\mathbb{E}(\hat{f}_n^2) \xi_n^2 / (\mathbb{E} \xi_n^2) + O(n^{-2}) . \]

A simple calculation shows that

(2.25) \[ \mathbb{E}(\hat{f}_n^2) \xi_n^2 = B/n + o(1/n) \]

with \( B \) given by (2.12). Combining (2.18)-(2.25) gives (1.35), finishing the proof.

COROLLARY. Jackknifing works: (1.38) holds.

THEOREM 5. Formula (1.40) holds.

Proof. By the converging-together lemma, it suffices to prove that

(2.26) \[ \sqrt{n} (r_{2n} - \bar{r}_{2n}) + 0 \]

in probability. Straightforward algebra shows that the left side of (2.26) equals

(2.27) \[ \sqrt{n} B_n (C_n - D_n) / (C_n + D_n) \]

where

\[ B_n = r(0,n - 1) - r(n,2n - 1) \]
\[ C_n = a(0,n - 1) / n \]
\[ D_n = a(n,2n - 1) / n \]
\[ a(a,b) = \sum_{j=a}^{b} a_j . \]

An argument the proof of (1.27) shows that

(2.28) \[ \sqrt{n} B_n \Rightarrow \sqrt{2} \alpha W(0,1)/\mathbb{E} \] .
Clearly

(2.29) \[ C_n - D_n + 0 \quad \text{a.s.} \]

Combining (2.27) through (2.29) verifies (2.26).
3. RESULTS AND PROOFS: II.

Theorem 6. Assume that (1.14), (1.17), and (1.18) hold under \( \hat{\varphi} \). Then, there exists a probability \( \hat{\varphi} \) such that:

1) \[ \frac{1}{t} \int_0^t \hat{\varphi}(X \circ \theta_s \circ \tau_s) \, ds + \hat{\varphi}(X \circ \tau_s) \text{ as } t \to \infty \]

2) \( \hat{\varphi}(X \circ \theta_u \circ \tau_s) = \hat{\varphi}(X \circ \tau_s) \) for \( u > 0 \)

3) \( \hat{\varphi}(X \circ \tau_s) = \frac{1}{S_1} \int_0^{S_1} \hat{\varphi}(X \circ \theta_s \circ \tau_s; \theta_s \circ \tau_s) \, ds \).

Proof. Let \( f \) be a bounded nonnegative function on \( \Omega \). For \( X \in \Omega \), set \( f_u(X) = f(X \circ \theta_u) \) and put

\[ q_u(X) = \int_0^{S_1} f_u(X \circ \theta_s \circ \tau_s) \, ds = q_0(X \circ \theta_u) \]

Proceeding as in the proof of Theorem 1, Birkhoff's ergodic theorem proves that

\[ \hat{P} = \frac{1}{n} \sum_{k=0}^{n-1} q_u(X \circ \theta_s \circ \tau_s) = \frac{1}{n} \int f(X \circ \theta_s \circ \tau_s) \, ds + E_q(X) \]

\( \hat{P} \text{ a.s., as } n \to \infty. \) The "squeeze" argument of (2.4), applied to (3.1), can be readily adapted to show that

\[ \frac{1}{n} \int_{S_u}^{s+u} f(X \circ \theta_s \circ \tau_s) \, ds + \frac{1}{S_1} E_q(X \circ \theta_u) \hat{P} \text{ a.s.} \]

It follows, from bounded convergence, that

\[ \frac{1}{t} \int_0^t \hat{\varphi}(X \circ \theta_s \circ \tau_s) \, ds + \frac{1}{S_1} E_q(X \circ \theta_u) \hat{P} \text{ a.s.} \]

as \( t \to \infty; \) in particular, specializing \( f \) to indicator functions yields
as \( t \to \infty \). Setting \( u = 0 \) in (3.2) proves i) and iii). For ii), observe that the left-hand side of (3.2) is independent of \( u \).

Throughout the remainder of this section, we assume that (1.14), (1.17), and (1.18) are in force for \( \hat{P} \), and that \( \mathbb{P} \) denotes probability under the measure \( \mathbb{P} \) constructed in Theorem 6.

**THEOREM 7.**

(3.3) \[ \mathbb{P}(r(t) + r) = 1. \]

**Proof.** Use definitions and then the fact that (1.24) holds under \( \hat{P} \) to get

(3.4) \[ \hat{P}(r(t) \hat{=} s + r; S_1 > s) = \hat{P}(\int_s^{s+t} f(X(u))du/t + r; S_1 > s) = \hat{P}(S_1 > s). \]

Now integrate formula (3.4) with respect to \( s \) from 0 to \( \infty \). On the right we get \( ES_1 \), positive by our assumptions. On the left, use dominated convergence to take the limit with respect to \( t \) inside the integral and then (1.13) to get \( \hat{P}(r(t) + r)ES_1 \). Cancelling \( ES_1 \) from both sides gives (3.3).

**THEOREM 8.** Under \( \mathbb{P} \) formulas (1.45) and (1.46) hold.

**Proof.** For any \( x, \)
By Billingsley ([1968], Theorem 20.2) the conditional probability above converges to the appropriate normal probability as $t \to \infty$, for every $x$ and $\theta$. Integrate both sides of (3.5) with respect to $s$ from $0$ to $\infty$, use (1.13) to see that the left side equals $P[\sqrt{t} (r(t) - r) < x] \mathbb{E}_\theta$, use dominated convergence on the right to take the limit with respect to $t$ inside the outer integral, and recall that any nonnegative random variable $Z$ has expectation $\int_0^\infty P[Z > t] dt$. Cancel $\mathbb{E}_\theta$ from both sides of (3.4) integrated as above to get that $P[\sqrt{t} (r(t) - r) < s]$ converges to the appropriate normal probability. A routine calculation verifies (1.46).
4. INDIRECT ESTIMATION OF TRANSACTION AVERAGES.

Suppose a performance measure \( s \) aggregates costs for transactions moving through the system. Let transaction \( i \) arrive at time \( A_i \) and leave at time \( B_i \). Put \( D_i = B_i - A_i \). Assume that the cost associated with transaction \( i \) is \( g_i(D_i) \) and that the average cost

\[
(4.1) \quad s = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} g_i(D_i)}{n}
\]

exists a.s.; clearly

\[
(4.2) \quad s_n = \frac{\sum_{i=1}^{n} g_i(D_i)}{n}
\]

consistently estimates \( s \). Our discussion carries over to more general transaction averages. We use \( s \) for concreteness.

In the notation of Section 1, state \( \tilde{X}_k \) includes the value of \( A_i \) for every transaction \( i \) in the system at time \( T_k \) and

\[
(4.3) \quad f(X(t)) = \sum_{i} g_i(t - A_i)
\]

where the sum is over those transactions \( i \) for which \( A_i < t < B_i \). Here \( g_i' \) is the right derivative of \( g_i \) and by definition \( g_i'(d) = 0 \) for \( d < 0 \) and for \( d > D_i \). This setup allows rather general \( g_i \)'s, though not indicators.

Now call the arrival rate \( \lambda \). Assuming all limits exist, \( r = \lambda \mu \) by Heyman and Stidham (1980). Usually \( \lambda \) is simply the reciprocal of the expected arrival spacing. In more general cases such as batch arrivals with batch size and spacing dependent, the theory developed earlier in this paper shows how to estimate \( \lambda \).

We can choose whether to observe discretized observations \( Y_n(f) \) as detailed earlier or transaction observations \( g_i(D_i) \). Our choice is based on four criteria:

(i) Ease of data collection. Formula (4.3) indicates that gathering discretized observations at state-change epochs is more work, but probably not much more because the evolution of the system has to be simulated in any case. Gathering discretized observations \( f(X(t)) \) at (reasonably short) equally-spaced intervals that do not necessarily correspond to state-change epochs may be much more work.

(ii) True variance. Some believe that \( \lambda s_n \) always has smaller variance than \( \bar{r}_m \) where \( n \) is the number of transactions completely processed up to event epoch \( m \). To some
extent the results of Carson and Law (1980) for special cases support this folklore, but Cooper's [1981], pp. 293-296 example shows that it is untrustworthy.

(iii) Variance estimation. Folklore has it that the sequence of transaction observations \( \{y_i(t)\} \) is covariance stationary if and only if the sequence of discretized observations \( \{y_n(t)\} \) is covariance stationary. Thus, loosely speaking, any variance-estimation technique (e.g., via batch means, spectral analysis, autoregressive representations, functional limit theorems) applies to both or to neither. In practice, however, one has to ask at what sample sizes asymptotic results reasonably apply: the faster covariance falls off with increasing lag, the better. Generally, for discretized observations, covariance does drop fairly quickly. By contrast, especially in systems that are not first-in, first-out (FIFO), a series of transaction observations scrambles past, present, and future and cuts connections between time spacing and index spacing. Thus, covariance between widely-spaced observations may well be significant. This messy covariance structure is hard to handle, at least with practical sample sizes.

(iv) Bias. The scrambling effect mentioned in point (iii) above may make it harder to detect (and hence attenuate) initialization bias for transaction observations. Incompletely processed transactions cause termination bias for transaction observations, exacerbated by the "inspection paradox" discussed earlier. For discretized observations, \( r_n \) is generally biased but our jackknife using \( \hat{r}_n \) makes this bias nearly negligible.

We conclude that, at least for non-FIFO systems, indirect estimation via discretized observations has more advantages than disadvantages relative to direct estimation via transaction observations. The results developed in the preceding sections make statistical analysis of the former possible and practical.

Asymptotically, \( s = r/\lambda \). Estimate \( s \) by plugging in consistent estimators for \( r \) and \( \lambda \). If, as is usual, \( \lambda \) is known, then divide confidence limits for \( r \) by \( \lambda \) to get confidence limits for \( s \). If \( \lambda \) has to be estimated, a straightforward ratio-estimator analog of the methods discussed in this paper must be used to obtain an asymptotically exact confidence interval.
REFERENCES


RLF/PWG/ed
Title: Estimating Time Averages via Randomly-Spaced Observations

Authors: Bennett L. Fox and Peter W. Glynn

Performing Organization: Mathematics Research Center, University of Wisconsin

Contract or Grant Number: DAAG29-80-C-0041

Type of Report & Period Covered: Summary Report - no specific reporting period

Prepared for: U.S. Army Research Office

Date: January 1984

Number of Pages: 20

Distribution Statement: Approved for public release; distribution unlimited.

Security Classification: UNCLASSIFIED

Abstract:
To estimate continuous-time averages via randomly-spaced observations of discrete-event systems, we develop a point-process framework and use it to generalize both regenerative and stationary-process oriented simulation methodologies. We give consistent estimators, central limit theorems, and an effective bias-reducing jackknife. The impact on indirect estimation of transaction (customer) averages is discussed.