ESTIMATING CUSTOMER AND TIME AVERAGES

PETER W. GLYNN

Stanford University, Stanford, California

BENJAMIN MELAMED

NEC Research USA, Inc., Princeton, New Jersey

WARD WHITT

AT&T Bell Laboratories, Murray Hill, New Jersey

(Received September 1990; revisions received January, November 1991; accepted November 1991)

In this paper we establish a joint central limit theorem for customer and time averages by applying a martingale central limit theorem in a Markov framework. The limiting values of the two averages appear in the translation terms. This central limit theorem helps to construct confidence intervals for estimators and perform statistical tests. It thus helps determine which finite average is a more asymptotically efficient estimator of its limit. As a basis for testing for PASTA (Poisson arrivals see time averages), we determine the variance constant associated with the central limit theorem for the difference between the two averages when PASTA holds.

This paper is concerned with estimating limiting time averages, customer (embedded) averages and their differences. When the two limiting averages are known to agree, that is, when arrivals see time averages (ASTA) (see Melamed and Whitt 1990a and references cited there), we want to know which finite average is a more asymptotically efficient estimator (i.e., produces smaller confidence intervals with large samples). When the two limiting averages need not agree, we want to estimate their difference and be able to test for ASTA.

To illustrate what happens when ASTA is known to hold, we give two examples.

Example 1: The Workload in the M/M/1 Queue

Let $\{U(t): t \geqslant 0\}$ be the continuous-time workload (or virtual waiting time) process in an M/M/1 queue with service rate 1 and arrival rate $\rho < 1$. Let $\{N(t): t \geqslant 0\}$ be the Poisson arrival process with associated arrival times $\{T_n : n \geqslant 1\}$. Then $\{U(T_n -) : n \geqslant 1\}$ is the sequence of waiting times (before beginning service). It is well known that the time average

$$V(t) = \frac{1}{t}\int_0^t U(s)\,ds, \quad t > 0, \tag{1}$$

and the customer average

$$W(t) = \frac{1}{N(t)}\sum_{k=1}^{N(t)} U(T_k -), \quad N(t) > 0, \tag{2}$$

both converge with probability one (w.p.1) to $\rho/(1-\rho)$ as $t \to \infty$, so that we have ASTA.

There is no need to estimate the limit in this case, but it is interesting to ask which estimator tends to produce smaller confidence intervals. We can easily decide, because it is known that central limit theorems (CLTs) hold, i.e.,

$$t^{1/2}[V(t) - v] \Rightarrow N(0, \sigma_v^2) \tag{3}$$

and

$$t^{1/2}[W(t) - w] \Rightarrow N(0, \sigma_w^2) \tag{4}$$

as $t \to \infty$, where $v = w = \rho/(1-\rho)$, $\Rightarrow$ denotes convergence in distribution and $N(m, \sigma^2)$ denotes a normally distributed random variable with mean m and variance σ^2. Moreover,

$$\sigma_v^2 = \frac{2\rho(3-\rho)}{(1-\rho)^4} \quad \text{and}$$

$$\sigma_w^2 = \frac{(2 + 5\rho - 4\rho^2 + \rho^3)}{(1-\rho)^4}, \tag{5}$$

(see Table II of Whitt 1989), so that

$$\sigma_w^2 - \sigma_v^2 = \frac{(2-\rho)(1+\rho)}{(1-\rho)^3} > 0. \tag{6}$$

Hence, for this example, the time average $V(t)$ is always asymptotically more efficient. (This result may seem to contradict results for the M/G/1 model by

Subject classifications. Queues, limit theorems: central limit theorems for customer and time averages. Queues, statistical inference: estimating customer and time averages. Statistics, estimation: averages over time and at embedded points.
Area of review. STOCHASTIC PROCESSES AND THEIR APPLICATIONS.

0030-364X/93/4102-0400 $01.25

Law (1975); he showed that $W(t)$ is superior to several indirect estimators, but $V(t)$ was not an alternative considered there.)

Example 2: A Uniformized Continuous-Time Markov Chain

Let $\{X(t):t \geq 0\}$ be an arbitrary, irreducible, finite-state, continuous-time Markov chain, uniformized so that the time between successive transitions has an exponential distribution independent of the state. (From the general case, this is achieved by introducing extra transitions from states to themselves; e.g., see Keilson 1979.) Let $\{N(t):t \geq 0\}$ be the Poisson process counting the transition epochs T_n. Then $\{X(T_n -):n \geq 1\}$ is the embedded discrete-time Markov chain. Let f be an arbitrary function on the state space of the chain and let $U(t) = f[X(t)]$. It is well known that both $V(t)$ in (1) and $W(t)$ in (2), defined in terms of $U(t)$ and T_n here, converge w.p.1 as $t \to \infty$ to $Ef[X(\infty)]$, where $X(\infty)$ has the unique stationary distribution of $\{X(t):t \geq 0\}$. Moreover, (3) and (4) hold with $v = w = Ef[X(\infty)]$.

Expressions for the variance constants σ_v^2 and σ_w^2 have been given by Kemeny and Snell (1960, 1961) and Hordijk, Iglehart and Schassberger (1976); see also Glynn (1984) and Whitt (1992). Moreover, Theorem 4.2 of Hordijk, Iglehart and Schassberger implies that $\sigma_w^2 < \sigma_v^2$. (They do not work directly with the averages in (1) and (2) but the subsequences in which t is replaced by the random epoch of the nth visit to a fixed state, i.e., regeneration points. The comparison is fair because both processes have the same random number of transitions. The comparison carries over to (1) and (2); see Lemma 1 on p. 210 of Glynn and Whitt (1986) and Section 6 of Glynn and Whitt (1989). The comparison of (1) and (2) is fair because both processes have the same expected number of transitions.) Hence, unlike Example 1, here the customer average $W(t)$ is always asymptotically more efficient.

Estimating the averages is also important when ASTA is not known to hold. We may want to use a customer average as an approximation for a time average, or vice versa. Then, we might want to estimate the difference to see if the approximation is reasonable, and possibly make corrections. For example, estimating time averages can be computationally intensive in discrete-event simulation. Time averages can be estimated if we collect data at every event and use the time intervals between these events. However, the total number of events may be prohibitively large. An alternative procedure is to introduce extra independent events strictly for sampling. However, it is typically easier to collect data only at time points in a subset of the original events. We might routinely use such a potentially biased procedure to estimate time averages, and only occasionally do a more complicated estimation, as described here, in order to estimate the bias to see if it is serious and, if so, to make corrections.

As a theoretical basis for constructing confidence intervals for estimators of the two limiting averages and their difference, we establish supporting central limit theorems. In Section 1 we specify the model. In Section 2 we show that the desired CLT follows from a CLT for basic component processes. In Section 3 we introduce martingale structure and give a sufficient condition for the CLT in Section 2. In Section 4 we show that the conditions in Section 3 are satisfied in a certain Markov framework. For this purpose, we apply the martingale CLT on p. 339 of Ethier and Kurtz (1986). For additional background on CLTs for general processes, see Chapter 4 of Billingsley (1968), Section 9.2 of Lipster and Shiryayev (1989), and Section 7.7 of Durrett (1991).

In Section 5 we show how the martingale CLT in Section 4 can be applied to generalized semi-Markov processes (GSMPs). In Section 6 we consider the special case of continuous-time Markov chains. Finally, in Section 7 we draw conclusions.

This paper was partly motivated by Yao (1992). Yao noted that the martingale CLT applies to a martingale associated with PASTA introduced by Wolff (1982). However, this martingale is not the natural estimator for the difference of the two limits; see Propositions 5 and 6 below. The analysis here is also similar in spirit to the analysis related to indirect estimation via $L = \lambda W$; see Law (1975), Carson and Law (1980) and Glynn and Whitt (1986, 1987, 1988, 1989). A new paper that presents additional related results concerning comparisons of estimators is Glasserman (1991) .

1. THE MODEL

Consider a continuous-time stochastic process $X \equiv \{X(t):t \geq 0\}$ taking values in a general space, and a stochastic point process on the interval $[0, \infty)$, characterized by the counting process $N \equiv \{N(t):t \geq 0\}$ or, equivalently, the sequence of successive points $\{T_n:n \geq 1\}$; i.e., $N(t) = \sup\{n \geq 0:T_n \leq t\}$, $t \geq 0$, where $T_0 = 0$ without there being a 0th point. Because of the many applications in queueing theory, we refer to N as the arrival process, even though the points need not be interpreted as arrivals. Let $U(t) = f(X(t))$, where f is a real-valued measurable function on the state space of X. We assume that the sample paths of U are left continuous with right limits, while the sample paths of N are right continuous with left limits.

We are interested in the time average $V(t)$ in (1) and the customer average $W(t)$ in (2), within the general framework here. Under appropriate regularity conditions,

$$V(t) \to v \quad \text{and} \quad W(t) \to w \quad \text{w.p.1 as } t \to \infty, \tag{7}$$

where the limits v and w satisfy the *bias formula*

$$w - v = \frac{\text{cov}[\mu^*(t), U^*(t)]}{E[\mu^*(t)]}, \tag{8}$$

with $U^*(t) = f(X^*(t))$, X^* and N^* being the stationary and stationary-increment versions of X and N, and $\mu^*(t)$ being the conditional intensity of a point from N^* at t conditional on $X^*(t)$; see Brémaud (1989), Makowski, Melamed and Whitt (1989), Melamed and Whitt (1990a, b), and Brémaud, Kannurpatti and Mazumdar (1991). See (17) below for an analog of (8) derived via the CLTs.

Moreover, under appropriate regularity conditions (see Sections 2–4), in this general framework, CLTs (3) and (4) hold, plus a joint CLT for $(W(t), V(t))$, i.e.,

$$t^{1/2}(V(t) - v, W(t) - w) \Rightarrow N(0, \Sigma) \quad \text{as } t \to \infty, \tag{9}$$

where the covariance matrix Σ has diagonal elements (variances) σ_v^2 and σ_w^2 and off-diagonal elements covariances σ_{vw}. As a consequence,

$$t^{1/2}(W(t) - v(t) - (w - v)) \Rightarrow N(0, \sigma_d^2) \quad \text{as } t \to \infty, \tag{10}$$

where

$$\sigma_d^2 = \sigma_w^2 + \sigma_v^2 - 2\sigma_{vw}, \tag{11}$$

Note that the limiting averages v and w appear in the CLTs in the translation terms.

In applications we will thus want to determine the variance constants σ_v^2, σ_w^2, σ_{vw} and σ_d^2 in (3), (4) and (9)–(11). In some cases, we will be able to calculate the variance constants analytically, but usually we must estimate them. For example, we might estimate σ_w^2 using batch means, i.e.,

$$\hat{\sigma}_w^2 = \frac{1}{n-1} \sum_{k=1}^{n} (n[W(kt/n) - W((k-1)t/n)] - W(t))^2.$$

See Goldsman and Meketon (1986), Glynn and Iglehart (1988), and Damerdji (1989a, b) and references cited there for more information about variance estimators.

2. A GENERAL CENTRAL LIMIT THEOREM

We first give general conditions for the vector process $[V(t), W(t)]$ and the difference $W(t) - V(t)$ to obey CLTs.

Proposition 1. a. *If there exist constants* v, w, *and* λ *such that as* $t \to \infty$

$$\{t^{-1/2}\left(\int_0^t U(s)\,ds - vt, \int_0^t U(s)\,N(ds) - \lambda wt, N(t) - \lambda t\right) \Rightarrow (L_1, L_2, L_3), \tag{12}$$

then

$$t^{1/2}(V(t) - v, W(t) - w) \Rightarrow (L_1, \lambda^{-1}(L_2 - wL_3)) \quad \text{as } t \to \infty \tag{13}$$

and

$$t^{1/2}(W(t) - V(t) - (w - v)) \Rightarrow \lambda^{-1}L_2 - \lambda^{-1}wL_3 - L_1 \quad \text{as } t \to \infty. \tag{14}$$

b. *If, in addition,* (L_1, L_2, L_3) *is normally distributed with means* 0, *variances* σ_i^2 *and covariances* σ_{ij}, *then the limits in* (13) *and* (14) *are normally distributed with*

$$\sigma_v^2 = \sigma_1^2, \quad \sigma_w^2 = \lambda^{-2}(\sigma_2^2 - 2w\sigma_{23} + w^2\sigma_3^2), \quad \text{and}$$

$$\sigma_d^2 = \sigma_1^2 + \lambda^{-2}\sigma_2^2 + \lambda^{-2}w^2\sigma_3^2 - 2\lambda^{-1}\sigma_{12} + 2\lambda^{-1}w\sigma_{13} - 2\lambda^{-2}w\sigma_{23}.$$

Proof. Note that

$$W(t) = \frac{1}{\lambda t}\int_0^t U(s)\,N(ds) - \left(\frac{t}{N(t)}\right)\left(\frac{N(t) - \lambda t}{t}\right)\frac{1}{\lambda t}\int_0^t U(s)\,N(ds),$$

so that

$$t^{1/2}(W(t) - w) = \frac{1}{\lambda t^{1/2}}\left[\int_0^t U(s)N(ds) - \lambda wt\right] - \left(\frac{t}{N(t)}\right)\left(\frac{N(t) - \lambda t}{t^{1/2}}\right)\frac{1}{\lambda t}\int_0^t U(s)\,N(ds)$$
$$\Rightarrow \lambda^{-1}L_2 - (\lambda^{-1})(L_3)(\lambda^{-1}\lambda w)$$
$$= \lambda^{-1}(L_2 - wL_3) \quad \text{as } t \to \infty.$$

This convergence above jointly with the normalized version of $V(t)$ follows from the convergence in (12) jointly with

$$\left[\frac{t}{N(t)}, \frac{1}{\lambda t}\int_0^t U(s)\, N(ds)\right] \Rightarrow (\lambda^{-1}, w),$$

which holds by Theorem 4.4 of Billingsley and the continuous mapping theorem, Theorem 5.1 of Billingsley.

Remark 1. The asymptotic efficiency of the estimator $[V(t), W(t)]$ in (13) can be improved by using a linear control estimator if λ is known; see Section 8 of Glynn and Whitt (1989) and the references cited there. Moreover, if $v = w$, then we can further improve the asymptotic efficiency by the same method.

3. MARTINGALE STRUCTURE

We now proceed to exploit martingale structure; for background, see Brémaud (1981) and Ethier and Kurtz (1986). Let $\{\mathscr{F}_t : t \geq 0\}$ be the internal history or filtration of (X, N); i.e., $\mathscr{F}_t$ is the σ-field generated by $\{[X(s), N(s)] : 0 \leq s \leq t\}$. Recall that we assumed that $U(t)$ is left continuous, so that it is predictable. We also assume that the point process $N(t)$ has a stochastic intensity $\mu(t)$. (The CLTs below extend to point processes with more general compensators.)

We now give a sufficient condition in this framework for the condition of Proposition 1 to hold.

Proposition 2. *The CLT* (12) *holds if there exist constants v, w, and λ such that:*

$$t^{-1/2}\Bigg[\int_0^t U(s)\, ds - vt,\quad \int_0^t \mu(s)\, ds - \lambda t,$$
$$\int_0^t U(s)\mu(s)\, ds - \lambda w t,$$
$$\int_0^t U(s)[N(ds) - \mu(s)\, ds],\quad N(t) - \int_0^t \mu(s)\, ds\Bigg]$$
$$\Rightarrow (Z_1, \ldots, Z_5) \quad \text{as } t \to \infty, \tag{15}$$

in which case $L_1 = Z_1$, $L_2 = Z_3 + Z_4$ and $L_3 = Z_2 + Z_5$. In addition, the asymptotic normality condition of Proposition 1b *is satisfied if $(Y_1, \ldots, Y_5)$ is normally distributed.*

Proof. Apply the continuous mapping theorem with addition.

Since U is predictable and μ is the intensity of N, the fourth and fifth components on the left in (15) are martingales under the usual moment conditions. Hence, CLTs can be established for these components directly, as Yao (1992) did for the fourth component when $\mu(t) = \lambda$ w.p.1. (For the CLT applications, it suffices to have f bounded and $E[N(t)^2] < \infty$ for all t.) The other three processes are centered cumulative processes, which satisfy CLTs under extra conditions such as appropriate mixing or regenerative structure, e.g., see Theorem 20 of Billingsley (1968) and Glynn and Whitt (1987). We will establish the joint convergence of all five components in the presence of extra Markov structure by showing that the vector process is asymptotically equivalent to a martingale. This Markov structure applies to a large class of models, e.g., after adding supplementary variables (see Glynn 1989a).

Remark 2. Note that nothing is gained by going from Proposition 1 to Proposition 2 when PASTA holds, because then $\mu(t) = \lambda$ w.p.1 and $v = w$; then the second term on the left in Proposition 2 is identically 0, while the third term is λ times the first.

Remark 3. Note that if the CLT (15) holds, then the limiting averages v and w can be identified from the asymptotic behavior of $[U(t), \mu(t)]$, without directly considering $N(t)$. In particular, if (15) holds, then

$$\left[V(t),\ t^{-1}\int_0^t \mu(s)\, ds,\ t^{-1}\int_0^t U(s)\mu(s)\, ds\right]$$
$$\xrightarrow{p} [v, \lambda, \lambda w] \quad \text{as } t \to \infty, \tag{16}$$

where $\xrightarrow{p}$ denotes convergence in probability, so that

$$\frac{\int_0^t U(s)\mu(s)\, ds}{\int_0^t \mu(s)\, ds} \xrightarrow{p} w \quad \text{as } t \to \infty, \tag{17}$$

which "explains" (8) and complements Melamed and Whitt (1990b) and Brémaud (1989).

4. A MARKOV FRAMEWORK

In this section, we assume that the basic processes X and N are defined in terms of an underlying Markov process Y, which is assumed to have left-continuous sample paths. As regularity conditions, let the state space of Y be a complete separable metric space and let the sample paths be continuous from the left and have limits from the right. Let the filtration $\{\mathscr{F}_t : t \geq 0\}$ be the internal filtration of Y.

In particular, we assume that there are measurable functions g and h so that

$$X(t) = g(Y(t)), \quad \mu(t) = h(Y(t)) \quad \text{and}$$
$$U(t) = f(g(Y(t))), \tag{18}$$

where f is the real-valued function defined in Section 1 and μ is the stochastic intensity defined in Section 2. Cases in which the intensity of N is a function of Y occur naturally when the points of N are completely determined by jumps of Y (see Section 5 of Melamed and Whitt 1990a, (3.2) of Melamed and Whitt 1990b, and p. 597 of Serfozo 1989).

As a consequence, the three cumulative processes in Proposition 2 can be represented as

$$\int_0^t U(s)\,ds = \int_0^t k_1(Y(s))\,ds$$
$$\int_0^t \mu(s)\,ds = \int_0^t k_2(Y(s))\,ds \tag{19}$$
$$\int_0^t U(s)\mu(s)\,ds = \int_0^t k_3(Y(s))\,ds$$

with $k_1(y) = f(g(y))$, $k_2(y) = h(y)$ and $k_3(y) = f(g(y))h(y)$. We provide conditions under which these cumulative processes are asymptotically equivalent to martingales. Let A be the infinitesimal generator of Y and $\mathscr{D}(A)$ its domain. We assume that there exist constants γ_i and functions b_i defined on the state space of Y and belonging to $\mathscr{D}(A)$ satisfying *Poisson's equation*, i.e.,

$$Ab_i = -k_i + \gamma_i \tag{20}$$

for $i = 1, 2$ and 3. Typically

$$\gamma_i = \pi k_i \equiv \int k_i(x)\,\pi(dx), \tag{21}$$

where π is the unique invariant probability measure of Y. (Typically the Markov process will be irreducible and positive recurrent.) See Glynn (1984, 1989b, c) and Whitt (1992) for further discussion.

Remark 4. In this setting, the limiting averages v, λ, and λw in (16) coincide with the steady-state means γ_1, γ_2, and γ_3 obtained via (19)–(21).

Since $b_i \in \mathscr{D}(A)$,

$$M_i(t) \equiv b_i(Y(t)) - \int_0^t (Ab_i)(Y(s))\,ds \tag{22}$$

is a martingale with respect to the filtration of Y; see p. 162 of Ethier and Kurtz.

Proposition 3. *If* (20) *holds and* $t^{-1/2}b_i(Y(t)) \Rightarrow 0$ *as* $t \to \infty$ *for* $i = 1, 2, 3$, *then*

$$t^{-1/2}\left(\int_0^t k_i(Y(s))\,ds - \gamma_i t - M_i(t)\right) \Rightarrow 0 \quad \text{as } t \to \infty,$$

where M_i *is the martingale in* (22).

Proof. If (20) holds, then

$$\int_0^t k_i(Y(s))\,ds - \gamma_i t$$
$$= -\int_0^t (Ab_i)(Y(s))\,ds$$
$$= -b_i(Y(t)) + b_i(Y(t)) - \int_0^t (Ab_i)(Y(s))\,ds.$$

Apply (22) and the condition.

Note that if Y is assumed to be stationary, then one condition in Proposition 3 is automatically satisfied, i.e.,

$$t^{-1/2}b_i(Y(t)) \stackrel{d}{=} t^{-1/2}b_i(Y(0)) \Rightarrow 0 \quad \text{as } t \to \infty, \tag{23}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Hence, to establish the CLT (15), it remains to establish the condition of Proposition 3 and the conditions of a martingale CLT. We apply the martingale CLT on p. 339 of Ethier and Kurtz. For additional background on CLTs for general processes, see Chapter 4 of Billingsley (1968), Section 9.2 of Lipster and Shiryayev (1989) and Section 7.7 of Durrett (1991). In particular, by the convergence-together theorem (Theorem 4.1 of Billingsley (1968) or Corollary 3.3 on p. 110 of Ethier and Kurtz), it suffices to apply the martingale CLT to the five-dimensional process $\mathbf{M}(t) = [M_1(t), \ldots, M_5(t)]$ having the first three components in (22), with (19), and the last two components in (15), i.e.,

$$M_4(t) = \int_0^t U(s)[N(ds) - \mu(s)\,ds] \quad \text{and}$$

$$M_5(t) = N(t) - \int_0^t \mu(s)\,ds, \quad t \geq 0. \tag{24}$$

Hence, we are in a position to apply the martingale CLT. We state the martingale CLT for $\mathbf{M}$ as a condition in our proposition. Sufficient conditions appear on p. 340 of Ethier and Kurtz.

Proposition 4. *If the conditions of Proposition* 3 *hold and* $t^{-1/2}\mathbf{M}(t) \Rightarrow \mathbf{Z}'$ *for* $\mathbf{M}$ *in* (22) *and* (24), *then the CLT* (15) *holds and* $Z_i = Z_i'$, $1 \leq i \leq 5$.

Under the regularity conditions given on p. 340 of Ethier and Kurtz, $t^{-1/2}\mathbf{M}(t)$ is asymptotically normally distributed as $t \to \infty$ with covariance elements determined by the quadratic variation processes associated with $\mathbf{M}$. In particular, suppose that $\mathbf{M}$ is square integrable, i.e., $E[|\mathbf{M}(t)|^2] < \infty$. Then we have

$$\frac{[M_i, M_i](t)}{t} \xrightarrow{p} \sigma_i^2 \quad \text{as } t \to \infty \tag{25}$$

and

$$\frac{[M_i, M_j](t)}{t} \xrightarrow{p} \sigma_{ij} \quad \text{as } t \to \infty, \tag{26}$$

where $[M_i, M_i] \equiv [M_i]$ is the *quadratic variation* and $[M_i, M_j]$ is the *cross variation* (see pp. 67, 79 of Ethier and Kurtz).

Remark 5. Another convenient view of the variance and covariance terms for $1 \leq i \leq 3$ comes from considering a stationary process framework, which is achieved by letting $Y(0)$ have the unique invariant probability measure π. Then

$$\sigma_{ij} = 2 \int_0^\infty \text{Cov}[k_i(Y(0)), k_j(Y(t))]\, dt;$$

e.g., see Theorem 20.1 of Billingsley (1968) and Section 9.2 of Lipster and Shiryayev (1989).

Remark 6. The CLT on p. 339 of Ethier and Kurtz is expressed in a stronger functional CLT form. There, a sequence of martingale processes $\{\mathbf{M}_n : n \geq 1\}$ is defined with $\mathbf{M}_n(t) = \mathbf{M}(nt)/\sqrt{n}$, $t \geq 0$. The associated limit process is then five-dimensional Brownian motion. We obtain the standard limit in R^5 referred to above by applying the continuous mapping theorem with the projection map at time $t = 1$.

From (24)–(26), it is easy to identify the limits for $4 \leq i, j \leq 5$. For this purpose, consider the typical case in which

$$U(T_k) \Rightarrow W \quad \text{and} \quad E[U(T_k)^2] \to E[W^2] \quad \text{as } k \to \infty, \tag{27}$$

so that

$$E[U(T_k)] \to E[W] = w \quad \text{as } k \to \infty. \tag{28}$$

Then

$$\frac{N(t)}{t} \xrightarrow{p} \sigma_5^2 = \lambda$$

$$\frac{1}{t} \sum_{k=1}^{N(t)} U(T_k) \xrightarrow{p} \sigma_{45} = \lambda E[W] = \lambda w \tag{29}$$

$$\frac{1}{t} \sum_{k=1}^{N(t)} U(T_k)^2 \xrightarrow{p} \sigma_4^2 = \lambda E[W^2] \quad \text{as } t \to \infty.$$

Moreover, if f is an indicator function, then $U(T_k)^2 = U(T_k)$ and $\sigma_4^2 = \sigma_{45} = \lambda w$.

Our next result is a CLT for the difference $W(t) - V(t)$ when PASTA holds. Equivalently, we assume that the stochastic intensity $\mu(t)$ is constant (so that N must be Poisson).

Proposition 5. *In addition to the conditions of Proposition* 4, *suppose that* (7) *holds, f is bounded, $E[N(t)^2] < \infty$ and $\mu(t) = \lambda$ w.p.*1 *for all t. Then the CLT* (14) *holds with $v = w$ and*

$$\sigma_d^2 = \lambda^{-1}\text{Var}[W]. \tag{30}$$

Proof. ASTA holds because (7) holds, f is bounded and $\mu(t) = \lambda$ w.p.1; e.g., see Theorems 1 and 2 of Melamed and Whitt (1990b). Consequently, $v = w$, $Z_2 = 0$ and $Z_3 = \lambda Z_1$; see Remark 2. The limit in (14) is $\lambda^{-1}L_2 - \lambda^{-1}wL_3 - L_1$, which coincides with $\lambda^{-1}Z_4 - \lambda^{-1}wZ_5$ under the extra assumptions here. Thus, by (29),

$$\sigma_d^2 = \lambda^{-2}\sigma_4^2 + \lambda^{-1}w^2 - 2\lambda^{-2}w\sigma_{45} = \lambda^{-1}\text{Var}[W].$$

Example 1 Revisited. As a consequence of Proposition 5, for the M/M/1 queue with arrival rate equal to the traffic intensity ρ discussed at the outset,

$$\sigma_d^2 = \lambda^{-1}\text{Var}[W] = \frac{2 - \rho}{(1 - \rho)^2}. \tag{31}$$

Note that σ_d^2 in (31) is asymptotically negligible compared to σ_v^2, σ_w^2 and $\sigma_w^2 - \sigma_v^2$ in (5) and (6) as $\rho \to 1$. From (11) and (31),

$$\sigma_{vw} = \frac{\rho(32 - 26\rho + 5\rho^2)}{2(1 - \rho)^4} > 0. \tag{32}$$

As a consequence of Proposition 5, when f is an indicator function, we obtain Yao's CLT for $M_4(t)$ plus a special form for the CLT for the difference.

Proposition 6. *In addition to the conditions of Proposition* 5, *suppose that f is an indicator function. Then*

$$t^{1/2}M_4(t) \Rightarrow N(0, \lambda w) \quad \text{as } t \to \infty \tag{33}$$

and (14) *holds with*

$$\sigma_d^2 = \lambda^{-1}w(1 - w). \tag{34}$$

Proof. Apply (29) and (30).

As a simple check of (33) and (34), let f be identically 1. Then obviously $W(t) = V(t) = 1$ for all t, so that $\sigma_d^2 = 0$.

Note that the relative standard error, defined by $\sqrt{\sigma_d^2}/w$, is

$$\frac{\sqrt{\sigma_d^2}}{w} = \sqrt{\frac{\lambda^{-1}(1 - w)}{w}}, \tag{35}$$

which becomes large as w gets small. Hence, the difference of small probabilities are hard to estimate in the sense of (35). An example would be small blocking probabilities in a loss model.

5. GENERALIZED SEMI-MARKOV PROCESSES

In applications to queues and related stochastic models, the underlying Markov process Y often will be associated with a generalized semi-Markov process (GSMP) (see Glynn 1989a). To treat this case, suppose that the sample paths of Y are piecewise linear with jumps occurring at the nondifferentiable points of the sample paths. Let $\{S_n : n \geq 1\}$ be the jump times of Y and let $J(t)$ count the number of jumps of Y in $[0, t]$. We assume that all the jump times of N are also jump times for Y, i.e., $\{T_n : n \geq 1\}$ is a subsequence of $\{S_n : n \geq 1\}$. Then the five quadratic variation processes and the ten cross-variation processes associated with **M** are:

$$[M_i, M_i](t) = \sum_{k=1}^{J(t)} [b_i(Y(S_k +)) - b_i(Y(S_k))]^2, \quad i = 1, 2, 3,$$

$$[M_4, M_4](t) = \sum_{k=1}^{N(t)} k_1(Y(T_k))^2,$$

$$[M_5, M_5](t) = N(t),$$

$$[M_i, M_j](t) = \sum_{k=1}^{J(t)} [b_i(Y(S_k+)) - b_i(Y(S_k))] \cdot [b_j(Y(S_k+)) - b_j(Y(S_k))], \quad 1 \leq i, j \leq 3, \ i \neq j,$$

$$[M_i, M_4](t) = \sum_{k=1}^{N(t)} [b_i(Y(T_k+)) - b_i(Y(T_k))]k_1(Y(T_k)), \quad 1 \leq i \leq 3,$$

$$[M_i, M_5](t) = \sum_{k=1}^{N(t)} [b_i(Y(T_k+)) - b_i(Y(T_k))], \quad 1 \leq i \leq 3,$$

$$[M_4, M_5](t) = \sum_{k=1}^{N(t)} k_1(Y(T_k)). \tag{36}$$

What we need, then, are weak laws of large numbers (WLLN) as specified in (25) and (26) for the quadratic and cross-variation processes in (36). Of course, we always have (29).

6. CONTINUOUS-TIME MARKOV CHAINS

As a concrete application of Section 5, we now suppose that Y is an irreducible, finite-state, continuous-time Markov chain (CTMC). Then A is the infinitesimal generator matrix and Poisson's equation (20) has a solution for γ_i defined by (21) for each i; see Section 4 of Whitt (1992). Then all solutions of (20) are of the form

$$b_i = Zk_i + (\pi b_i)e, \tag{37}$$

where π is the unique invariant measure, e is a column vector of 1's and Z is the fundamental matrix with components

$$Z_{lm} = \int_0^\infty [P_{lm}(t) - \pi_m] \, dt, \tag{38}$$

with $P(t)$ being the time-dependent transition matrix associated with A (see (13) and Proposition 4.2 of Whitt 1992). In this setting the conditions for the CLT are easy to verify.

Proposition 7. *If Y is an irreducible finite-state CTMC, then the conditions of Proposition* 4 *are satisfied so that the CLT* (15) *holds. The limit has a multivariate normal distribution with covariance matrix elements given by* (15), (26), (29) *and* (36).

Given that $Y(S_k) = l$ and $Y(S_k+) = m$,

$$b_i(Y(S_k+)) - b_i(Y(S_k)) = (Zk_i)_l - (Zk_i)_m = \sum_j (Z_{lj} - Z_{mj})k_i(j). \tag{39}$$

From (36) and (39), we thus deduce familiar expressions for σ_i^2 and σ_{ij} for $1 \leq i, j \leq 3$; e.g., see Remark 5 above and (12) of Whitt (1992). For ways to calculate σ_i^2 and σ_{ij}, see Proposition 4.2 and its Corollary 3 in Whitt (1992).

Proposition 8. *For the finite-state CTMC and* $1 \leq i, j \leq 3$,

$$\sigma_i^2 = 2 \sum_l \sum_m k_i(l)\pi_l Z_{lm} k_i(m)$$

$$\sigma_{ij} = 2 \sum_l \sum_m k_i(l)\pi_l Z_{lm} k_j(m).$$

Proof. We only derive the expression for σ_i^2; the derivation for σ_{ij} is similar. Let A_l be the sum of the off-diagonal elements of the lth row of A. Since A_{lj}/A_l is the transition probability of the embedded chain in state l at epochs S_k,

$$P(Y(S_k) = l) \to \pi_l A_l \Big/ \sum_j \pi_j A_j \quad \text{as } k \to \infty$$

and

$$\frac{J(t)}{t} \to \sum_j \pi_j A_j \text{ w.p.1} \quad \text{as } t \to \infty.$$

Hence, by (36) and (39),

$$\sigma_i^2 = \sum_j \pi_j A_j \sum_l \sum_m \left(\frac{\pi_l A_l}{\sum_j \pi_j A_j}\right)\left(\frac{A_{lj}}{A_l}\right)[(Zk_i)_l - (Zk_i)_m]^2$$

$$= \sum_l \sum_m \pi_l A_{lm}[(Zk_i)_l - (Zk_i)_m]^2$$

$$= \sum_l \sum_m \pi_l A_{lm}[(Zk_i)_l^2 - 2(Zk_i)_l(Zk_i)_m + (Zk_i)_m^2]$$

$$= -2 \sum_l \sum_m \pi_l A_{lm}(Zk_i)_l(Zk_i)_m \quad \text{(explanation below)}$$

$$= -2 \sum_l \sum_m \sum_j \sum_n \pi_l A_{lm} Z_{lj} k_i(j) Z_{mn} k_i(n)$$

$$= 2 \sum_l \sum_j k_i(l) \pi_l Z_{lj} k_i(j) \quad \text{(explanation below).}$$

The fourth equality above holds because $\pi A = 0$ and $Ae = 0$. The final equality holds because

$$\sum_m \sum_n A_{lm} Z_{mn} k_i(n) = -k_i(l),$$

which holds by (20) and (37).

By similar reasoning, we can also obtain expressions for σ_{i4} and σ_{i5}, $1 \leq i \leq 3$, if we make some additional assumptions. Suppose that the subsequence $\{T_k\}$ corresponds to all transitions from l to m for a set of pairs (l, m). Then $\{Y(T_k)\}$ is a discrete-time Markov chain (DTMC) with stationary distribution say π^*.

Proposition 9. *If $\{T_k\}$ and π^* are defined as above and $t^{-1}N(t) \to \lambda$ as $t \to \infty$, then*

$$\sigma_{i4} = \lambda \sum_l \sum_m \pi_l^* \left(\frac{A_{lm}}{A_l}\right)[(Zk_i)_l - (Zk_i)_m] U(l)$$

and

$$\sigma_{i5} = \lambda \sum_l \sum_m \pi_\lambda^* \left(\frac{A_{lm}}{A_l}\right)[(Zk_i)_l - (Zk_i)_m]$$

for $1 \leq i \leq 3$.

7. CONCLUSION

We have provided a general framework for establishing central limit theorems for time and customer averages. The value of this approach is demonstrated by the explicit formula for the difference variance constant when PASTA holds in (30). The explicit expressions for the quadratic variation and cross variation processes in a GSMP framework in (36) should also assist further study.

ACKNOWLEDGMENT

We thank David Yao for sending us his paper, which was a source of motivation.

REFERENCES

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

BRÉMAUD, P. 1981. *Point Processes and Queues.* Springer-Verlag, New York.

BRÉMAUD, P. 1989. Characteristics of Queueing Systems Observed at Events and the Connection Between Stochastic Intensity and Palm Probability. *Queue. Syst.* **5,** 99–112.

BRÉMAUD, P., R. KANNURPATTI AND R. MAZUMDAR. 1991. Event and Time Averages: A Review and Some Generalizations. *Adv. Appl. Prob.* (to appear).

CARSON, J. S., AND A. M. LAW. 1980. Conservation Equations and Variance Reduction in Queueing Simulations. *Opns. Res.* **28,** 535–546.

DAMERDJI, H. 1989a. Strong Consistency of the Variance Estimator in Steady-State Simulation Output Analysis. Department of Operations Research, Stanford University, Stanford, Calif.

DAMERDJI, H. 1989b. Strong Consistency of the Spectral Variance Estimator. Department of Operations Research, Stanford University, Stanford, Calif.

DURRETT, R. 1991. *Probability: Theory and Examples.* Wadsworth and Brooks/Cole, Pacific Grove, Calif.

ETHIER, S. N., AND T. G. KURTZ. 1986. *Markov Processes, Characterization and Convergence.* John Wiley, New York.

GLASSERMAN, P. 1991. Filtered Monte Carlo. Graduate School of Business, Columbia University, New York.

GLYNN, P. W. 1984. Some Asymptotic Formulas for Markov Chains With Applications to Simulation. *J. Statist. Comput. Simul.* **19,** 97–112.

GLYNN, P. W. 1989a. A GSMP Formalism for Discrete Event Systems. *Proc. IEEE* **77,** 14–23.

GLYNN, P. W. 1989b. Poisson's Equation for the Recurrent M/G/1 Queue. Department of Operations Research, Stanford University, Stanford, Calif.

GLYNN, P. W. 1989c. A Lyapunov Bound for Solutions of Poisson's Equation. Department of Operations Research, Stanford University, Stanford, Calif.

GLYNN, P. W., AND D. L. IGLEHART. 1988. A New Class of Strongly Consistent Variance Estimators for Steady-State Simulations. *Stoch. Proc. Appl.* **28,** 71–80.

GLYNN, P. W., AND W. WHITT. 1986. A Central Limit Theorem Version of $L = \lambda W$. *Queue. Syst.* **1,** 191–215.

GLYNN, P. W., AND W. WHITT. 1987. Sufficient Conditions for Functional-Limit-Theorem Versions of $L = \lambda W$. *Queue. Syst.* **1,** 279–287.

GLYNN, P. W., AND W. WHITT. 1988. Ordinary CLT and WLLN Versions of $L = \lambda W$. *Math. Opns. Res.* **13,** 674–692.

GLYNN, P. W., AND W. WHITT. 1989. Indirect Estimation Via $L = \lambda W$. *Opns. Res.* **37,** 82–103.

GOLDSMAN, D., AND M. MEKETON. 1986. A Comparison of Several Variance Estimators for Stationary Increment Stochastic Processes. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Ga.

HORDIJK, A., D. L. IGLEHART AND R. SCHASSBERGER. 1976. Discrete Time Methods for Simulating Continuous Time Markov Chains. *Adv. Appl. Prob.* **8,** 772–788.

KEILSON, J. 1979. *Markov Chain Models—Rarity and Exponentiality.* Springer-Verlag, New York.

KEMENY, J. G., AND J. L. SNELL. 1960. *Finite Markov Chains.* Van Nostrand, Princeton, N.J.

KEMENY, J. G., AND J. L. SNELL. 1961. Finite Continuous Time Markov Chains. *Theor. Prob. Appl.* **6,** 101–105.

LAW, A. M. 1975. Efficient Estimators for Simulated Queueing Systems. *Mgmt. Sci.* **22,** 30–41.

LIPSTER, R. SH., AND A. N. SHIRYAYEV. 1989. *Theory of Martingales.* Kluwer, Boston.

MAKOWSKI, A., B. MELAMED AND W. WHITT. 1989. On Averages Seen by Arrivals in Discrete Time. In *Proceedings 28th IEEE Conference on Decision and Control,* IEEE, New York, 1084–1086.

MELAMED, B., AND W. WHITT. 1990a. On Arrivals That See Time Averages. *Opns. Res.* **38,** 156–172.

MELAMED, B., AND W. WHITT. 1990b. On Arrivals That See Time Averages: A Martingale Approach. *J. Appl. Prob.* **27,** 276–384.

SERFOZO, R. F. 1989. Poisson Functionals of Markov Processes and Queueing Networks. *Adv. Appl. Prob.* **21,** 595–611.

WHITT, W. 1989. Planning Queueing Simulations. *Mgmt. Sci.* **35,** 1341–1366.

WHITT, W. 1992. Asymptotic Formulas for Markov Processes With Applications to Simulation. *Opns. Res.* **40,** 279–291.

WOLFF, R. W. 1982. Poisson Arrivals See Time Averages. *Opns. Res.* **30,** 223–231.

YAO, D. D. 1992. On Wolff's PASTA Martingale. *Opns. Res.* **40,** S352–S355.