

Parameter Estimation for Partially Observed Queues

Thomas M. Chen, *Member, IEEE*, Jean Walrand, *Fellow, IEEE*, and David G. Messerschmitt, *Fellow, IEEE*

Abstract—In this paper, we consider parameter estimation for a FIFO queue with deterministic service times and two independent arrival streams of “observed” and “unobserved” packets. The arrivals of unobserved packets are Poisson with an unknown rate λ while the arrivals of observed packets are arbitrary. Maximum likelihood estimation of λ is formulated based on the arrival times and waiting times of k observed packets. The likelihood function is derived in terms of the transition probabilities of the unfinished work process which are calculated recursively. Sufficient conditions for consistency, asymptotic normality, and asymptotic efficiency are given. The mean and variance of the MLE are measured in simulation experiments. Numerical results indicate that the MLE is consistent and asymptotically normal.

I. INTRODUCTION

IN PRACTICE, it is sometimes difficult to directly observe an important aspect of behavior of a queueing system. For example, an automatic teller machine may record only the service times of each customer; it is unable to observe directly the queue length or customers’ waiting times, which are the principal performance measures of the teller service, so instead, these performances measures must be statistically inferred from the available data (using an assumed system model). In such situations, the unknown system model parameters may be the arrival rate or service rate or both, and estimation of them leads to an inference about the performance measure of interest. In other cases, the available data may be the queue length or waiting times, and the goal is estimation of the unknown system model parameters [1]. For instance, a subscriber to a packet switching network might inject a stream of packets and make an estimate of the network load based on their observed round-trip delays.

Previous studies of parameter estimation have been mainly restricted to the $M/M/1$ queue [2]–[4]; assumed complete and continuous observations of the queueing system [2]–[5]; or depended on approximations [6]. In this paper, we consider parameter estimation for a partially observed queue with a known deterministic service rate and two independent arrival streams of “observed” and “unobserved” packets. The arrivals of unobserved packets are Poisson with an unknown rate λ while the arrivals of observed packets may be arbitrary. It is assumed that arrival times, waiting times, and departure times are known for observed packets only.

Paper approved by I. Chlamtac, the Editor for Computer Networks of the IEEE Communications Society. Manuscript received July 15, 1991; revised March 31, 1992. This work was supported under a grant from Pacific Bell and a matching grant from the University of California MICRO program.

T. M. Chen is with GTE Laboratories, Inc., Waltham, MA 02254 USA.

J. Walrand and D. G. Messerschmitt are with the Department of Electrical Engineering and Computer Science, Berkeley, CA 94720 USA.

IEEE Log Number 9400574.

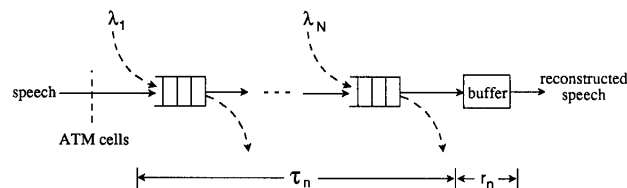


Fig. 1. ATM switched speech.

Maximum likelihood estimation of λ is formulated based on the data for k observed packets.

In Section II, the likelihood function is derived in terms of the transition probabilities of the unfinished work process; these transition probabilities are calculated recursively. Sufficient conditions are given for consistency, asymptotic normality, and asymptotic efficiency of the maximum likelihood estimate (MLE). Section III presents simulation results indicating that the log-likelihood function is continuous, concave, unimodal, and approximately quadratic over short intervals. The results also indicate that the MLE is consistent and asymptotically normal.

Applications of this estimation problem may exist in network control situations where the state of network queues cannot be continuously observed. Another possible application, if this problem is generalized to multiple queues in tandem, could be in ATM switched speech as shown in Fig. 1. ATM (asynchronous transfer mode) is a standardized fast packet switching technique based on 53 byte cells (fixed-length packets) for the Broadband Integrated Services Digital Network [7]. At the speaker, encoded speech is packetized into a stream of ATM cells which are injected into the network. The cells travel through a number of queues in tandem and experience random sojourn times $\{\tau_n\}$ through the network due to contention with other network traffic (shown as dotted lines).

At the listener, the cells are stored in a buffer where additional delays $\{r_n\}$ are imposed such that the total delay

$$\tau_n + r_n = D \quad (1)$$

for every cell is the same constant D , thereby compensating for the variability of the $\{\tau_n\}$, before the speech information is “played out” (decoded) [8]. Those cells with sojourn times $\tau_n > D$ will exceed their “playout” times and are effectively lost. Performance measures are the end-to-end delay D and the probability of cell loss, P_{loss} . Ideally, both D and P_{loss} should be minimized but there is a direct trade-off (in general, a smaller D corresponds to a larger P_{loss}).

In the process of selecting the target value of D , it is important to estimate the (unknown) tradeoff between D and P_{loss} , i.e., the cumulative probability distribution function of the $\{\tau_n\}$. At least, it is desirable to estimate the upper tail

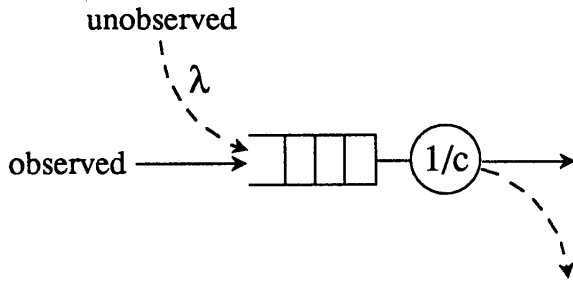


Fig. 2. Partially observed queue.

distribution of the $\{\tau_n\}$. There are many possible approaches to estimating the distribution depending on the stochastic model assumed for the $\{\tau_n\}$ (e.g., see [9]). One possible approach is to model the actual queueing processes, for example, by assuming the interfering traffic at each queue are Poisson. Clearly, this would be an approximation to the actual arrival process but the Poisson approximation may be justifiable when the traffic is the aggregation of a large number (e.g., more than 10) of independent streams [10]–[12]. We may estimate the unknown arrival rates (by the method of maximum likelihood estimation, for instance) and then estimate the steady-state probability distribution of the $\{\tau_n\}$ by queueing theory analysis. The target value of D may then be chosen, say, as the p -percentile of the $\{\tau_n\}$.

II. MAXIMUM LIKELIHOOD ESTIMATION

A. The Model

Fig. 2 represents a FIFO queue with deterministic service times c and two independent arrival streams of packets. The arrivals of unobserved packets are Poisson with an unknown rate λ while the arrivals of observed packets may be arbitrary. A hypothetical observer positioned at the output of the queue is able to monitor the departure times $\{d_n\}$ and sojourn times $\{\tau_n\}$ of the observed packets only (sojourn times are the sum of waiting times and service times, and might be recorded by timestamping [13]). Given d_n and τ_n , the n th arrival time is simply $a_n = d_n - \tau_n$, and the waiting time in the queue is $w_n = \tau_n - c$. Thus, any of (d_n, τ_n) , (a_n, τ_n) , or (a_n, d_n) can be regarded as a complete data record for the n th observed packet. Unobserved packets cannot be monitored by the hypothetical observer. However, their arrival rate λ clearly has a statistical effect on the observed sojourn times $\{\tau_n\}$ because all packets must contend for service (in general, a larger λ would imply larger $\{\tau_n\}$). We consider maximum likelihood estimation of λ based on the observations $\{(a_n, \tau_n)\}$ for k observed packets.

B. The Likelihood Function

The likelihood function is the *a priori* probability of observing the $\{(a_n, \tau_n)\}$ as a function of the unknown parameter λ . First note that the $\{a_n\}$ are independent of λ and thus may be regarded as constants in the likelihood function. The $\{\tau_n\}$ are samples of the unfinished work in the system, denoted by $U(t)$, which is defined as the amount of time required

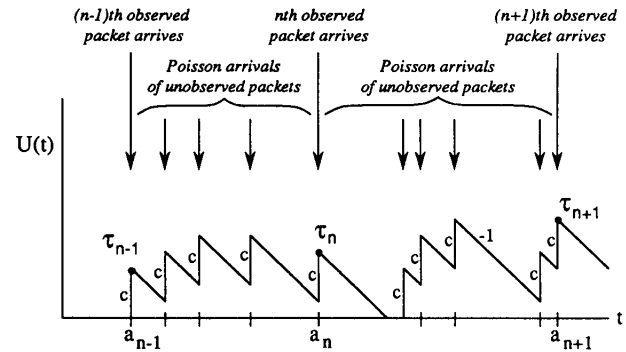


Fig. 3. A realization of the unfinished work in the system.

to complete service for all the packets in the system at time t . The unfinished work is also called virtual waiting time [14], [15] because a hypothetical packet arriving at time t would experience a waiting time of $U(t)$ in the queue. It is a continuous-time, right-continuous process with nonnegative real values. As shown in Fig. 3, upward instantaneous jumps of magnitude c occur at each arrival time. Otherwise, $U(t)$ decreases with a slope of -1 until it reaches zero where it remains until the next arrival.

A few remarks should be made about $U(t)$. First, note that τ_n is the value of $U(t)$ sampled at the time $t = a_n$. Thus, the observations imply $U(a_n) = \tau_n$ and $U(a_n-) = \lim_{t \uparrow a_n} U(t) = \tau_n - c$. Second, only the jump times of $U(t)$ are random; otherwise, the behavior of $U(t)$ is deterministic. Hence, $U(a_n) = \tau_n$ depends only on $U(a_{n-1}) = \tau_{n-1}$ and the arrival times of unobserved packets in the time interval (a_{n-1}, a_n) which are Poisson with rate λ . Third, the sequence $\{\tau_n\}$ is a Markov process.

It may be seen then that the likelihood of the observations $\{(a_1, \tau_1), \dots, (a_k, \tau_k)\}$ can be written as the product

$$L_k(\lambda) = \prod_{n=2}^k p_n(\tau_n | \tau_{n-1}; \lambda) \quad (2)$$

where $p_n(x | y; \lambda)$ is the probability density function of $U(a_n)$ conditional on $U(a_{n-1}) = y$. Strictly speaking, $L_k(\lambda)$ is a *conditional* likelihood function because it neglects the probability of the first observation τ_1 . However, the probability of the first observation can be ignored since it becomes unimportant as $k \rightarrow \infty$, and we shall refer to $L_k(\lambda)$ as the likelihood function with the understanding that there is little difference for asymptotic results.

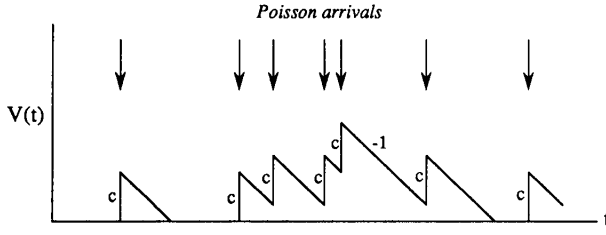
The difficulty is the determination of the transition probabilities of $U(t)$. Define a process $V(t)$ as shown in Fig. 4; the jump times of $V(t)$ are Poisson with rate λ , and otherwise $V(t)$ decreases with a slope of -1 until it reaches zero. Let $F(x, t | x_0; \lambda)$ be the transition distribution function of $V(t)$,

$$F(x, t | x_0; \lambda) \equiv \Pr \{V(t) \leq x | V(0) = x_0\}, \quad (3)$$

and $f(x, t | x_0; \lambda)$ be the transition density function.

Since $U(t)$ and $V(t)$ behave identically in any interval (a_{n-1}, a_n) conditional on $U(a_{n-1}) = V(a_{n-1})$, we can see that

$$p_n(x | y; \lambda) = f(x - c, a_n - a_{n-1} | y; \lambda). \quad (4)$$

Fig. 4. A realization of $V(t)$.

The likelihood function may then be written as

$$L_k(\lambda) = \prod_{n=2}^k f(\tau_n - c, a_n - a_{n-1} | \tau_{n-1}; \lambda). \quad (5)$$

In fact, this is not the exact likelihood function we will use because the density function $f(x, t | x_0; \lambda)$ may contain singularities, i.e., delta functions. It is easier to evaluate the likelihood function defined as

$$L_k(\lambda) = \prod_{n=2}^k \frac{1}{2\Delta} [F(\tau_n - c + \Delta, a_n - a_{n-1} | \tau_{n-1}; \lambda) - F(\tau_n - c - \Delta, a_n - a_{n-1} | \tau_{n-1}; \lambda)] \quad (6)$$

for some very small Δ , which can be derived by approximating $f(x, t | x_0; \lambda)$ with a central difference with approximation error of $O(\Delta^2)$. That is, we are making the approximation

$$f(x, t | x_0; \lambda) \approx \frac{1}{2\Delta} \Pr\{x - \Delta < V(t) \leq x + \Delta | V(0) = x_0\} \quad (7)$$

and the likelihood function (6) can be interpreted as the probability of a small volume around the observation point $(\tau_1, \dots, \tau_k)$. The value of Δ affects the likelihood function only by a constant scaling factor and does not affect its extrema if Δ is sufficiently small, e.g., $\Delta \ll c$. Maximization of the likelihood function (6) avoids the difficulty with delta functions in $f(x, t | x_0; \lambda)$. The maximum-likelihood estimate (MLE) $\hat{\lambda}_k$ will be defined as the value of λ maximizing the likelihood function (6) or the log-likelihood function $l_k(\lambda) = \log L_k(\lambda)$.

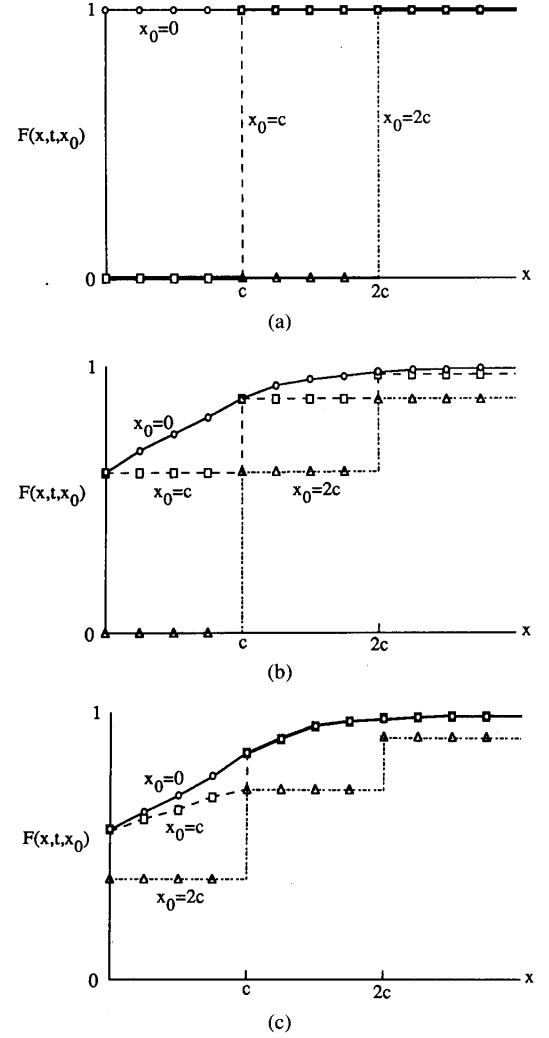
The final step consists of finding the transition distribution function $F(x, t | x_0; \lambda)$. It is known to satisfy the Takacs integrodifferential equation [14], [15]:

$$\frac{\partial}{\partial t} F(x, t | x_0; \lambda) = \frac{\partial}{\partial x} F(x, t | x_0; \lambda) - \lambda F(x, t | x_0; \lambda) + \lambda F(x - c, t | x_0; \lambda) I\{x \geq c\} \quad (8)$$

where $I\{\cdot\}$ is the indicator function assuming the value 1 when its argument is true. However, the solution has not been known except in terms of Laplace transforms which are difficult to invert. One of our main results is a recursion formula for $F(x, t | x_0; \lambda)$:

Proposition 1: The transition distribution function is

$$F(x, t | x_0; \lambda) = \sum_{i=0}^{u(x+t+x_0)} \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

Fig. 5. $F(x, t | x_0; \lambda)$ as a function of x for (a) $t = 0$, (b) $t = c$, and (c) $t = 2c$.

$$- \sum_{i=u(x+c)}^{u(x+t+x_0)} \frac{\lambda^i (ic - x)^i}{i!} e^{-\lambda(ic-x)} F(0, x+t-ic | x_0; \lambda) \quad (9)$$

for $t, x, x_0 \geq 0$ where $u(y) \equiv \lfloor y/c \rfloor$, and $\lfloor y \rfloor$ denotes the greatest integer less than or equal to y .

Proof: See Appendix A. $\square$

Examples of $F(x, t | x_0; \lambda)$ for $\lambda c = 0.5$ are shown in Figs. 5 and 6. It can be seen that $F(x, t | x_0; \lambda)$ would be difficult to obtain from inverse transforms because it is irregular and discontinuous in x and t .

Unfortunately, it is difficult to maximize the log-likelihood function $l_k(\lambda)$ directly, and we must resort to numerical optimization methods [16], [17]. Numerical methods search for the maximum over a range of λ by evaluating the log-likelihood function at various points (hopefully with the fewest number of evaluations as possible). Numerical optimization methods are effective if the function is sufficiently well-behaved, e.g., continuous, unimodal, and concave. These characteristics of the log-likelihood function are investigated in Section III. Note that each evaluation of the log-likelihood function involves the

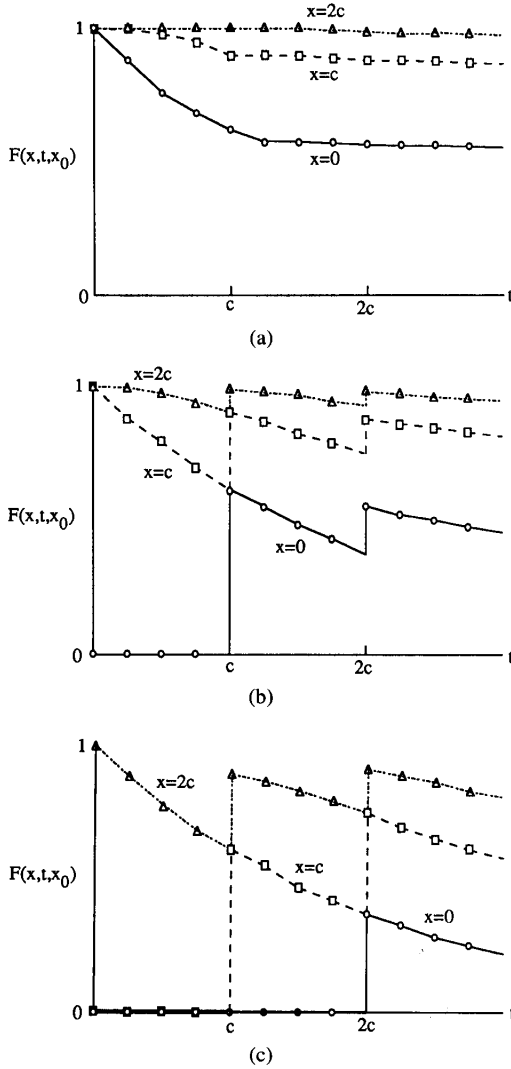


Fig. 6. $F(x, t | x_0; \lambda)$ as a function of t for (a) $x_0 = 0$, (b) $x_0 = c$, and (c) $x_0 = 2c$.

calculation of $k - 1$ terms, hence the complexity of parameter estimation performed for a particular k is linearly proportional to k . For sequential estimation performed for multiple values of k , however, this complexity might be reduced by storing the evaluations of the log-likelihood function involved in obtaining $\hat{\lambda}_k$; then when τ_{k+1} is observed, these evaluations might be used to save some amount of calculations for finding $\hat{\lambda}_{k+1}$.

C. Asymptotic Properties of the MLE

In the case of independent observations, it is well known that maximum likelihood estimates are consistent under certain “regularity” conditions, and asymptotically normal and efficient under additional conditions [18]–[20]. The point here, however, is that the observations $\{\tau_n\}$ are not independent. Sufficient conditions for these asymptotic properties may also be found for dependent (Markovian) observations (e.g., see [18]). The traditional approach is to show: (1) that the likelihood equation has a consistent root, (2) the asymptotic distribution of the “total score” at the true parameter is normal, and (3) then the higher terms of a Taylor expansion of the total

score around the true parameter become negligible for large samples with high probability. Most likely, these conditions are unnecessarily restrictive; the difficulty lies in finding the necessary and sufficient conditions, which is a challenging problem for future study.

Let λ_0 denote the true value of λ (we will use λ as a variable). Define the scores,

$$u_n(\lambda) = \frac{\partial}{\partial \lambda} \log p_n(\tau_n | \tau_{n-1}; \lambda), \quad (10)$$

the total score,

$$S_k(\lambda) = \sum_{n=2}^k u_n(\lambda), \quad (11)$$

and the Fisher information,

$$I_k(\lambda) = E[S_k^2(\lambda)]. \quad (12)$$

The MLE $\hat{\lambda}_k$ is the solution of the likelihood equation $S_k(\lambda) = 0$. A probability density function $p(x; \lambda)$ is called regular with respect to its m th λ -derivative if

$$\frac{\partial^m}{\partial \lambda^m} \int_{-\infty}^{\infty} p(x; \lambda) dx = \int_{-\infty}^{\infty} \frac{\partial^m}{\partial \lambda^m} p(x; \lambda) dx. \quad (13)$$

Roughly speaking, regularity means that there are no discontinuities of $p(x; \lambda)$ in x that are dependent on λ .

Consistency of the MLE is established first.

Theorem 1: Suppose for $n \geq 2$,

- i) $p_n(\tau_n | \tau_{n-1}; \lambda)$ is regular with respect to its first λ -derivative,
- ii) $E[u_n^2(\lambda_0)] < \infty$,
- iii) $u_n(\lambda)$ is continuous in λ for all values of τ_n .

If $\hat{\lambda}_k$ is the unique solution of the likelihood equation $S_k(\lambda) = 0$ for $k \geq$ some k_0 , then the MLE $\hat{\lambda}_k$ is a consistent estimator of λ_0 .

Proof: By Lemma 1 (in Appendix B) and a law of large numbers for martingales [21, p. 243],

$$\frac{1}{k-1} S_k(\lambda_0) \rightarrow 0 \quad (14)$$

with probability one as $k \rightarrow \infty$. With condition iii), it implies there exists a sequence of solutions of $S_k(\lambda) = 0$ that converges to λ_0 [20, p. 359]. If $\hat{\lambda}_k$ is the unique solution for $k \geq k_0$, then the sequence $\{\hat{\lambda}_k\}$ must converge to λ_0 . $\square$

The regularity condition i) and continuity condition iii) are a matter of verification from (9). Condition ii) is difficult to confirm but may be justified. It would be sufficient to establish that $|\frac{\partial}{\partial \lambda} \log f(x, t | x_0; \lambda)|$ is bounded for all admissible values of λ wherever $f(x, t | x_0; \lambda) > 0$. Assuming $|\frac{\partial}{\partial \lambda} f(x, t | x_0; \lambda)|$ is bounded, the only points where $|\frac{\partial}{\partial \lambda} \log f(x, t | x_0; \lambda)|$ is unbounded are those points where $f(x, t | x_0; \lambda) = 0$. The final condition, that the MLE is the unique maximum of the log-likelihood function, is indicated by the results in Section III.

Asymptotic normality of the MLE is a stronger property than consistency and requires additional assumptions. We will use the following proposition about the limiting normal distribution of the total score.

Proposition 2: Suppose, in addition to the conditions in Theorem 1, as $k \rightarrow \infty$,

- i) $\sum_{n=2}^k u_n^2(\lambda_0)/I_k(\lambda_0) \rightarrow 1$ in probability,
- ii) $E[\max_{2 \leq n \leq k} u_n^2(\lambda_0)]/I_k(\lambda_0) \rightarrow 0$,

then as $k \rightarrow \infty$,

$$S_k(\lambda_0)\sqrt{I_k(\lambda_0)} \rightarrow N(0, 1) \quad (15)$$

in distribution where $N(0, 1)$ denotes the standard normal distribution.

Proof: From Lemma 1, we have $I_k(\lambda_0) = \sum_{n=2}^k E[u_n^2(\lambda_0)]$. The final result follows directly from a martingale limit theorem [18, p. 385]. $\square$

The asymptotic normality of the MLE follows from a Taylor expansion of the total score around $\lambda = \lambda_0$.

Theorem 2: Suppose, in addition to the conditions in Proposition 2,

- i) $p_n(\tau_n | \tau_{n-1}; \lambda)$ is regular with respect to its second λ -derivative,
- ii) $S'_k(\lambda_0)/E[S'_k(\lambda_0)] \rightarrow 1$ in probability as $k \rightarrow \infty$,
- iii) $S''_k(\lambda)$ is bounded in neighborhood of $\lambda = \lambda_0$ by some integrable function $g(\tau_1, \dots, \tau_k)$,

then as $k \rightarrow \infty$,

$$\sqrt{I_k(\lambda_0)}(\hat{\lambda}_k - \lambda_0) \rightarrow N(0, 1). \quad (16)$$

Proof: The consistency of the MLE $\hat{\lambda}_k$ allows the Taylor expansion

$$S_k(\hat{\lambda}_k) = S_k(\lambda_0) + (\hat{\lambda}_k - \lambda_0)S'_k(\lambda_0) + \frac{1}{2}(\hat{\lambda}_k - \lambda_0)^2 S''_k(\lambda^*) \quad (17)$$

where $|\lambda^* - \lambda_0| < |\hat{\lambda}_k - \lambda_0|$. By definition $S_k(\hat{\lambda}_k) = 0$, and hence

$$\frac{S_k(\lambda_0)}{\sqrt{I_k(\lambda_0)}} = \sqrt{I_k(\lambda_0)}(\hat{\lambda}_k - \lambda_0) \cdot \left(\frac{-S'_k(\lambda_0) - \frac{1}{2}(\hat{\lambda}_k - \lambda_0)S''_k(\lambda^*)}{I_k(\lambda_0)} \right). \quad (18)$$

From Lemma 2 (in Appendix B) and conditions ii) and iii), the right-hand side converges in probability to $\sqrt{I_k(\lambda_0)}(\hat{\lambda}_k - \lambda_0)(1 + e_k)$ where $|e_k| \leq |\lambda^* - \lambda_0|g(\tau_1, \dots, \tau_k)/I_k(\lambda_0)$. Note λ^* is necessarily consistent, and thus $|e_k| \leq o(1)$ in probability [19, p. 294]. The final result follows after applying Proposition 2 on the left hand side of (18). $\square$

We observe that the variance of the asymptotic distribution of $\hat{\lambda}_k$ achieves the Cramer-Rao bound, and therefore:

Corollary 1: Under the conditions of Theorem 2, the MLE $\hat{\lambda}_k$ is asymptotically efficient.

It is rather difficult to verify the conditions for asymptotic normality and efficiency which essentially rely on some ergodic properties of the $\{u_n(\lambda_0)\}$. As the $\{u_n(\lambda_0)\}$ are random variables in the $(\tau_1, \dots, \tau_n)$ -space, it might be conjectured that the conditions hold if the $\{\tau_n\}$ are ergodic, for example, when the observed arrivals are periodic. However, the precise interpretation of the conditions in verifiable terms is not altogether clear. The simulation results in the next section indicate that the MLE is asymptotically normal when the observed arrivals are deterministic.

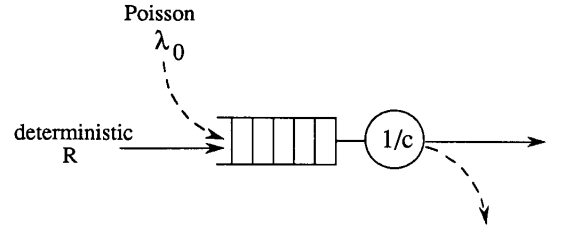


Fig. 7. Simulated queueing model.

III. SIMULATION RESULTS

A. The Likelihood Function

In this section, we investigate the log-likelihood function in simulation experiments. Another issue concerns the accuracy of $\hat{\lambda}_k$ as a function of k . Although asymptotic properties of the MLE can be established under certain conditions, they make no statement about the mean and variance of the MLE for small samples. Below we present some numerical results on the accuracy of the MLE as a function of k .

The queueing model shown in Fig. 7 was simulated. The arrivals of unobserved packets are Poisson with rate λ_0 while the arrivals of observed packets are deterministic with rate R . For convenience, the deterministic service rate was normalized to $1/c = 1$. We used the method of independent replications [22], [23]. On each simulation run, the observations $\{(a_n, \tau_n)\}$ for k observed packets were recorded after the queue was allowed to approach steady-state (with observed packets absent). Let M denote the number of replicated simulation runs. On the m th simulation run, let $l_{k,m}(\lambda)$ denote the log-likelihood function and $\hat{\lambda}_{k,m}$ be the MLE calculated from $l_{k,m}(\lambda)$. Since the replicated runs are identical and independent, $\{\hat{\lambda}_{k,1}, \dots, \hat{\lambda}_{k,M}\}$ are i.i.d. samples of the MLE $\hat{\lambda}_k$.

The mean of $\hat{\lambda}_k$ is estimated by the usual sample mean

$$\bar{\lambda}_k = \frac{1}{M} \sum_{m=1}^M \hat{\lambda}_{k,m} \quad (19)$$

and the variance of $\hat{\lambda}_k$ is estimated by the usual sample variance

$$s^2 = \frac{1}{M-1} \sum_{m=1}^M (\hat{\lambda}_{k,m} - \bar{\lambda}_k)^2. \quad (20)$$

The 95% confidence interval of the sample mean is

$$\bar{\lambda}_k \pm \frac{s}{\sqrt{M}} t_{M-1, .975} \quad (21)$$

and the 95% confidence interval of s^2 is

$$\left(\frac{(M-1)s^2}{\chi_{M-1, .975}^2}, \frac{(M-1)s^2}{\chi_{M-1, .025}^2} \right) \quad (22)$$

where $t_{M-1,p}$ is the p -percentile of the t -distribution with $M-1$ degrees of freedom, and $\chi_{M-1,p}^2$ is the p -percentile of the χ^2 -distribution with $M-1$ degrees of freedom.

Numerical optimization methods are effective if $l_k(\lambda)$ is sufficiently "well-behaved," i.e., continuous, unimodal, and concave [16], [17]. Even better, if $l_k(\lambda)$ is quadratic or nearly quadratic over λ , there are well-known efficient search

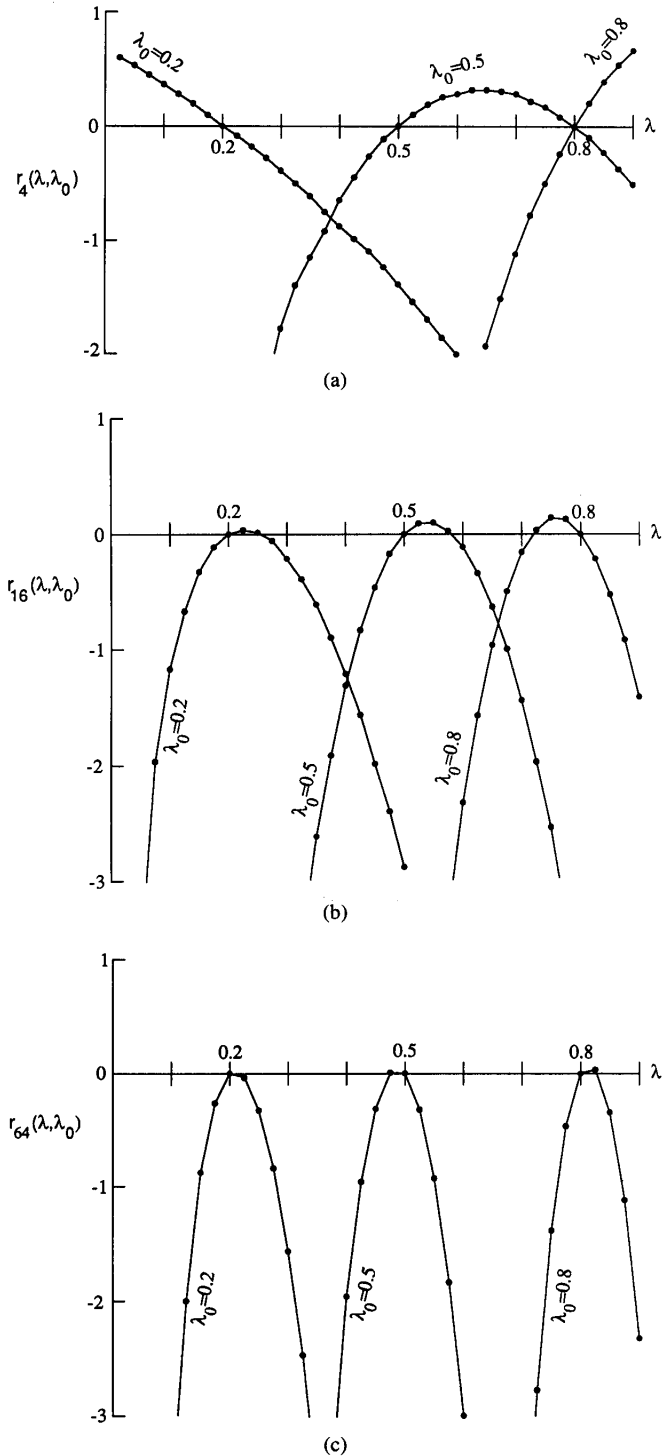


Fig. 8. Examples of $r_k(\lambda, \lambda_0)$ for $R = 0.1$, (a) $k = 4$, (b) $k = 16$, and (c) $k = 64$.

methods using quadratic interpolation. Such characteristics are difficult to prove analytically because $F(x, t | x_0; \lambda)$ is a complicated expression. These characteristics were investigated for many independent realizations of $l_k(\lambda)$ over a range of parameters. Examples are shown in Fig. 8 of the log-likelihood ratio

$$r_k(\lambda, \lambda_0) \equiv \log \left(\frac{L_k(\lambda)}{L_k(\lambda_0)} \right) = l_k(\lambda) - l_k(\lambda_0) \quad (23)$$

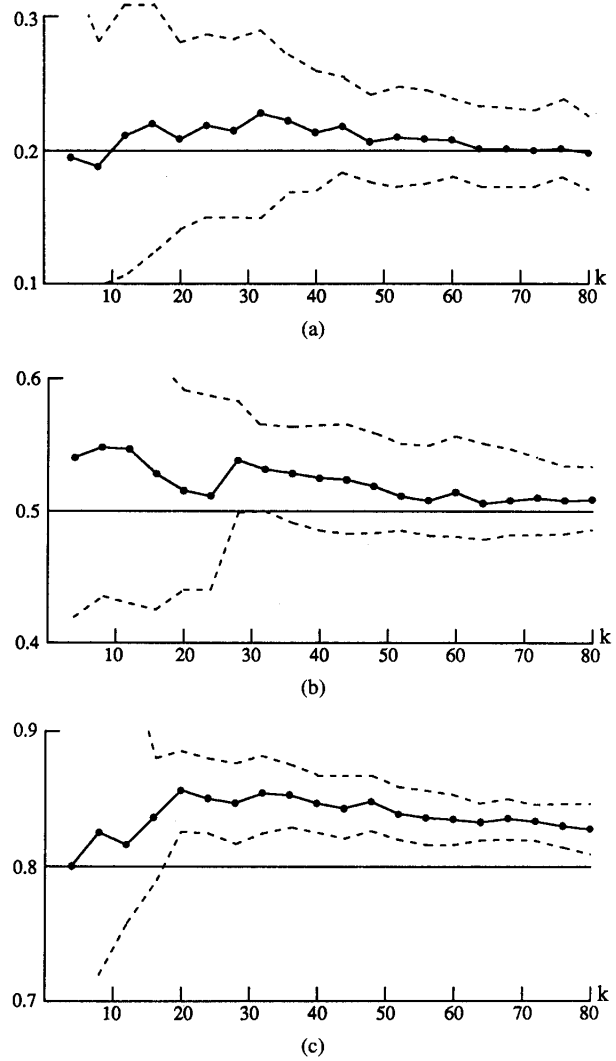


Fig. 9. Measured mean of the MLE $\hat{\lambda}_k$ with 95% confidence intervals for $R = 0.1$, (a) $\lambda_0 = 0.2$, (b) $\lambda_0 = 0.5$, and (c) $\lambda_0 = 0.8$.

which is simply the log-likelihood function normalized by its value at $\lambda = \lambda_0$ (we set $\Delta = 10^{-3}c$ but the exact value of Δ is unimportant). Fig. 8 is typical in the sense that all realizations of $l_k(\lambda)$ observed in the experiments were similarly found to be continuous, unimodal, and concave in λ . In addition, it was found that $l_k(\lambda)$ is approximately quadratic over short intervals of λ , and Powell's quadratic interpolation method [16], [17] performed well for finding the maximum of $l_k(\lambda)$.

B. Accuracy of the MLE

Fig. 8 indicates that $l_k(\lambda)$ is unimodal and continuous in λ , and hence we would expect that the MLE $\hat{\lambda}_k$ is a consistent estimator. In addition, since the observed arrivals are periodic, $\hat{\lambda}_k$ might be expected to be asymptotically normal and efficient.

The mean of $\hat{\lambda}_k$ measured with $M = 8$ is shown in Fig. 9 for $R = 0.1$. A slight bias, perhaps due to the effect of the initial observation, can be seen for small k . However, as expected, the bias diminishes as $k \rightarrow \infty$. The standard deviation of $\hat{\lambda}_k$ measured with $M = 8$ is plotted in Fig. 10. It can be seen

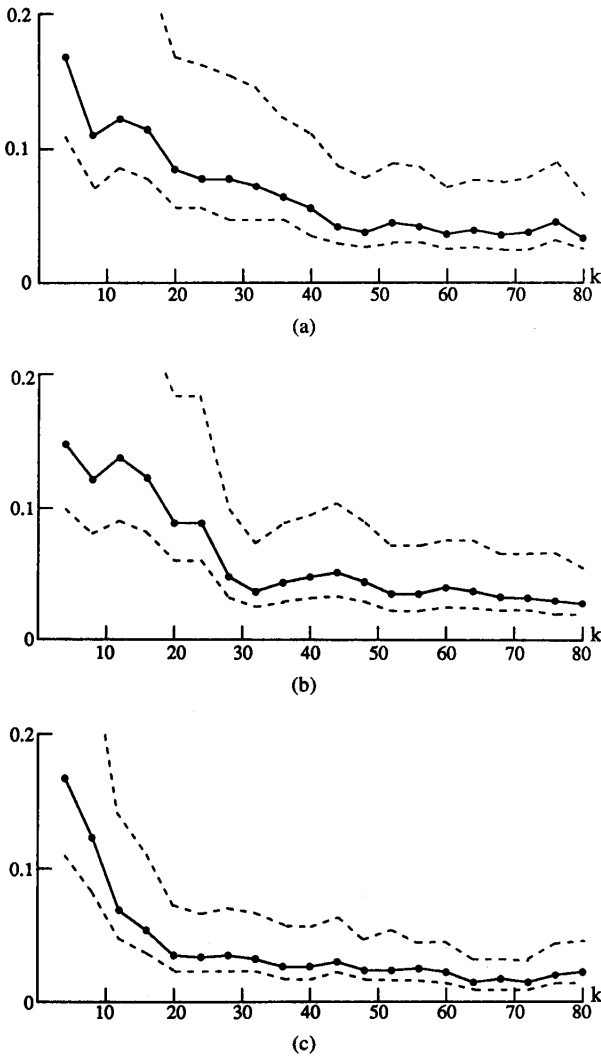


Fig. 10. Measured standard deviation of the MLE $\hat{\lambda}_k$ with 95% confidence intervals for $R = 0.1$, (a) $\lambda_0 = 0.2$, (b) $\lambda_0 = 0.5$, and (c) $\lambda_0 = 0.8$.

that the variance of $\hat{\lambda}_k$ diminishes with k , which implies that $\hat{\lambda}_k$ converges to its asymptotic mean. This agrees with the expectation that $\hat{\lambda}_k$ is consistent. Results for $R = 0.05$ were found to be very similar. In general, the results were not found to be very sensitive to the exact values of R or λ_0 .

In Figs. 11–13, the empirical density functions (histograms) of $\hat{\lambda}_k$ are shown for $k = 16$ and $k = 64$ (based on $M = 88$ samples). For $k = 16$, it is apparent that $\hat{\lambda}_k$ is not normally distributed. But for $k = 64$, the histogram of $\hat{\lambda}_k$ resembles a normal density. The usual χ^2 goodness-of-fit test was used to test the hypothesis that for $k = 64$, $\hat{\lambda}_k$ is normally distributed with the observed means and variances. A significance level of $\alpha = 0.05$ was chosen. For $\lambda_0 = 0.2$, the observed statistic $\chi_{obs}^2 = 3.62$ was much less than the critical value $\chi_{8,.95}^2 = 15.51$. For $\lambda_0 = 0.5$, the observed statistic $\chi_{obs}^2 = 1.20$ was also much less than the critical value $\chi_{5,.95}^2 = 11.07$. For $\lambda_0 = 0.8$, the observed statistic $\chi_{obs}^2 = 5.64$ was again less than the critical value $\chi_{4,.95}^2 = 9.49$. Thus, the test led to acceptance of the hypothesis, which supported our initial conjecture that the MLE $\hat{\lambda}_k$ is asymptotically normal. Results for $R = 0.05$ were found to be very similar.

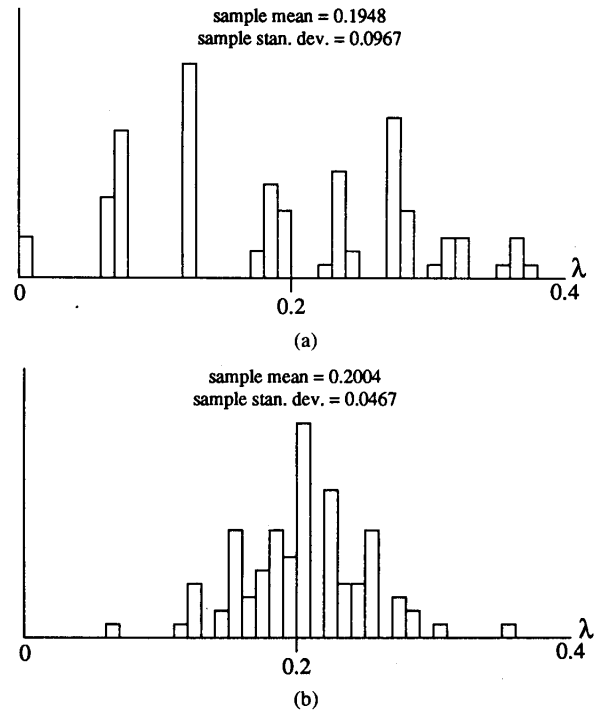


Fig. 11. Empirical probability density function (histograms) of the MLE $\hat{\lambda}_k$ for $R = 0.1$, $\lambda_0 = 0.2$, (a) $k = 16$, and (b) $k = 64$.

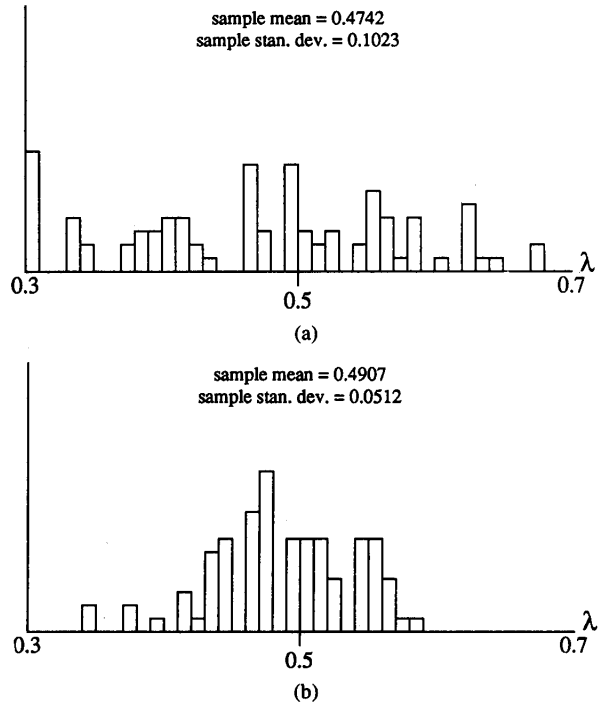


Fig. 12. Empirical probability density function (histograms) of the MLE $\hat{\lambda}_k$ for $R = 0.1$, $\lambda_0 = 0.5$, (a) $k = 16$, and (b) $k = 64$.

IV. CONCLUSIONS

This paper has considered parameter estimation for partially observed queues with a known deterministic service rate and unobserved Poisson traffic. Maximum likelihood estimation of the unobserved arrival rate was formulated in terms of the transition probabilities of the unfinished work process which are calculated recursively. Sufficient (but probably un-

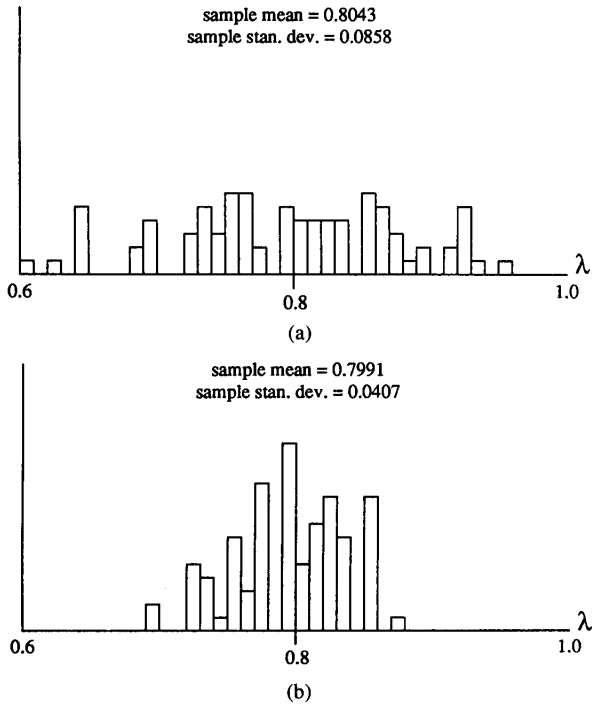


Fig. 13. Empirical probability density function (histograms) of the MLE $\hat{\lambda}_k$ for $R = 0.1$, $\lambda_0 = 0.8$, (a) $k = 16$, and (b) $k = 64$.

necessarily restrictive) conditions for consistency, asymptotic normality, and asymptotic efficiency were given. A question for future research is the necessary and sufficient conditions for these asymptotic properties. In simulation experiments, the log-likelihood function was found to be continuous, unimodal, concave, and approximately quadratic over short intervals. Powell's quadratic interpolation method performed well for maximizing the log-likelihood function. Numerical results indicated agreement with the consistency and asymptotically normality of the MLE. Another question for future research is the accuracy (mean and variance) of the MLE for any k .

While the discussion here has been restricted to the continuous-time context, obviously the same problem can also be formulated in discrete-time which might be more appropriate for some applications. The results in discrete-time may be expected to be entirely similar, which seems to be the case [24].

The maximum likelihood estimation method presented here shows promise for extensions to other queueing models, although the appropriate method for each statistical inference problem will depend on the available data. Possible applications may exist in network control situations where the state of a particular queue cannot be continuously observed. Such estimation problems have been hitherto largely neglected in classical queueing theory.

APPENDIX A

We take the same approach as the derivation of the Takacs equation and begin with the appropriate random walk process [14]. Instead of deriving the Takacs equation however, we will seek a recursive expression for the probability distribution function for the random walk and then take the limit as the

time increment Δt becomes very small. We approximate $V(t)$ by a random walk $X_n \approx V(n\Delta t)$ with a reflecting barrier at $x = 0$, where $\Delta t < c$. The meaning of reflecting barrier [14] here is not to be confused with reflecting barriers for diffusion processes. It means that whenever X_n goes to $x = 0$ or below, it stays at $x = 0$ until the next upward transition. The random walk proceeds as

$$X_{n+1} = (X_n + y_n)^+ \equiv \max(X_n + y_n, 0) \quad (\text{A.1})$$

where the $\{y_n\}$ are i.i.d., and $y_n = ic - \Delta t$ if there are i arrivals in a time increment. There are i arrivals in Δt with probability

$$f_i = e^{-\lambda\Delta t} \frac{(\lambda\Delta t)^i}{i!}, \quad i = 0, 1, 2, \dots \quad (\text{A.2})$$

and therefore

$$\Pr\{y_n = ic - \Delta t\} = f_i. \quad (\text{A.3})$$

It is known that the probability distribution function of X_n ,

$$H_n(x) \equiv \Pr\{X_n \leq x\} \quad (\text{A.4})$$

obeys the recurrence relationship

$$H_n(x) = \begin{cases} \int_0^\infty H_{n-1}(y)f(x-y)dy & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad (\text{A.5})$$

where $f(y)$ is the probability density function of the $\{y_n\}$. Notice that this is almost a convolution except that $H_n(x) = 0$ for $x < 0$. By substitution, the basic recurrence relationship becomes

$$H_n(z) = f_0 H_{n-1}(x + \Delta t) I\{x \geq 0\} + \sum_{i=1}^{\infty} f_i H_{n-1}(x + \Delta t - ic) I\{x \geq ic - \Delta t\}. \quad (\text{A.6})$$

We now use transforms to find $H_n(x)$ from any initial condition $H_0(x) = I\{x \geq x_0\}$. The Laplace transform

$$H_n^*(s) = \int_0^\infty e^{-sx} H_n(x) dx \quad (\text{A.7})$$

exists for $\text{Re}(s) > 0$ because $H_n(x) \leq 1$ (all x). The recurrence relationship in terms of transforms becomes

$$\begin{aligned} H_n^*(s) &= \int_0^\infty f_0 e^{sx} H_{n-1}(x + \Delta t) dx \\ &\quad + \sum_{i=1}^{\infty} \int_{ic-\Delta t}^\infty f_i e^{-sx} H_{n-1}(x + \Delta t - ic) dx \\ &= \sum_{i=0}^{\infty} \int_0^\infty f_i e^{-s(x+ic-\Delta t)} H_{n-1}(x) dx \\ &\quad - \int_0^{\Delta t} f_0 e^{-s(x-\Delta t)} H_{n-1}(x) dx \\ &= \exp\{\Delta t(s - \lambda + \lambda e^{-sc})\} H_{n-1}^*(s) \\ &\quad - \int_0^{\Delta t} f_0 e^{-s(x-\Delta t)} H_{n-1}(x) dx \\ &= \exp\{\Delta t(s - \lambda + \lambda e^{-sc})\} H_{n-1}^*(s) - \psi_{n-1}(s) \end{aligned} \quad (\text{A.8})$$

if $\Delta t < c$ where we define

$$\psi_n(s) \equiv \int_0^{\Delta t} f_0 e^{-s(x-\Delta t)} H_n(x) dx \quad (\text{A.9})$$

for convenience. Notice that the first term of $H_n^*(s)$ represents a convolution of $H_{n-1}(y)$ and $f(y)$, and the last term is a correction due to the barrier.

Repeating this recurrence from the initial condition $H_0^*(s) = s^{-1}e^{-sx_0}$ leads to

$$H_n^*(s) = \exp\{n\Delta t(s - \lambda + \lambda e^{-sc})\} s^{-1} e^{-sx_0} - \sum_{i=0}^{n-1} \exp\{i\Delta t(s - \lambda + \lambda e^{-sc})\} \psi_{n-i-1}(s). \quad (\text{A.10})$$

This can be inverted without difficulty for $x \geq 0$ to get

$$H_1(x) = \sum_{j=0}^{u_1} f_j(1) \quad (\text{A.11})$$

and

$$H_n(x) = \sum_{i=0}^{u_n} f_i(n) - \sum_{i=2}^n f_0(1) f_{z_i}(i-1) H_{n-i}(x + i\Delta t - z_i c) \cdot I\left\{z_i \leq \frac{x + i\Delta t}{c} < z_i + \frac{\Delta t}{c}\right\} \quad (\text{A.12})$$

for $n \geq 2$ where

$$z_i \equiv \lfloor (x + i\Delta t)/c \rfloor, \quad u_i \equiv \lfloor (x + i\Delta t - x_0)/c \rfloor, \\ f_i(n) \equiv \frac{(n\lambda\Delta t)^i}{i!} e^{-n\lambda\Delta t}. \quad (\text{A.13})$$

When Δt becomes very small, $H_n(x)$ should converge to $F(x, t | x_0; \lambda) = F(x, n\Delta t | x_0; \lambda)$. As $\Delta t \rightarrow 0$, note that the indicator function has nonzero values only when $(x + i\Delta t)/c$ has integer values from $\lfloor (x + c)/c \rfloor$ to $\lfloor (x + n\Delta t)/c \rfloor$. If $(x + i\Delta t)/c = j$ for some integer j , then $f_0(1) f_{z_i}(i-1)$ converges to $\frac{(j\rho - \lambda x)^j}{j!} e^{-(j\rho - \lambda x)}$ where $\rho \equiv \lambda c$. Also, $f_i(n)$ converges to $\frac{(\lambda t)^i}{i!} e^{-\lambda t}$. Thus for very small Δt , the equation for $H_n(x)$ can be simplified, leading to the recursion equation:

$$F(x, t | x_0; \lambda) = \sum_{i=0}^{u(x+t+x_0)} \frac{(\lambda t)^i}{i!} e^{-\lambda t} - \sum_{i=u(x+c)}^{u(x+t)} \frac{(i\rho - \lambda x)^i}{i!} e^{-(i\rho - \lambda x)} F(0, x + t - ic | x_0; \lambda) \quad (\text{A.14})$$

for $t > 0$ where $u(y) \equiv \lfloor y/c \rfloor$. Notice that $F(0, t | x_0; \lambda) = \Pr\{V(t) = 0 | V(0) = x_0\}$, and hence $F(0, t | x_0; \lambda) = 0$ for $t < x_0$. Thus, the upper bound on the second summation can be lowered to $u(x + t + x_0)$, leading to the final result.

APPENDIX B

Lemma 1: If $p_n(\tau_n | \tau_{n-1}; \lambda)$ is regular with respect to its first λ -derivative, then

$$E[u_n(\lambda_0)] = 0 \quad (\text{B.1})$$

and the $\{u_n(\lambda_0)\}$ are uncorrelated.

Proof:

$$E[u_n(\lambda_0) | \tau_{n-1}] = 0$$

is implied by the regularity assumption. (B.2)

$$E[u_n(\lambda_0)] = E[E\{u_n(\lambda_0) | \tau_{n-1}\}] = 0. \quad (\text{B.3})$$

$$E[u_{n+m}(\lambda_0)u_n(\lambda_0)] \\ = E[E\{u_{n+m}(\lambda_0)u_n(\lambda_0) | \tau_{n-1}\}], \quad m \neq 0 \\ = E[E\{E\{u_{n+m}(\lambda_0) | \tau_{n+m-1}\}u_n(\lambda_0) | \tau_{n-1}\}] \\ = 0. \quad (\text{B.4})$$

Lemma 2: If $p_n(\tau_n | \tau_{n-1}; \lambda)$ is regular with respect to its second λ -derivative, then

$$I_k(\lambda_0) = E[-S'_k(\lambda_0)]. \quad (\text{B.5})$$

Proof:

$$E[u_n^2(\lambda_0)] \\ = E[E\{u_n^2(\lambda_0) | \tau_{n-1}\}] \\ = E\left[\frac{\partial}{\partial \lambda} E\{u_n(\lambda) | \tau_{n-1}\} - E\left\{\frac{\partial}{\partial \lambda} u_n(\lambda) | \tau_{n-1}\right\}\right]_{\lambda=\lambda_0} \\ \text{implied by regularity} \\ = E\left[E\left\{-\frac{\partial}{\partial \lambda} u_n(\lambda) | \tau_{n-1}\right\}\right]_{\lambda=\lambda_0} \\ = E\left[-\frac{\partial}{\partial \lambda} u_n(\lambda)\right]_{\lambda=\lambda_0}. \quad (\text{B.6})$$

$$I_k(\lambda_0) = \sum_{n=2}^k E[u_n^2(\lambda_0)] = \sum_{n=2}^k E\left[-\frac{\partial}{\partial \lambda} u_n(\lambda)\right]_{\lambda=\lambda_0} \\ = E[-S'_k(\lambda_0)]. \quad (\text{B.7})$$

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their constructive suggestions and Prof. D. Daley for reading an earlier manuscript.

REFERENCES

- [1] D. Gross and C. Harris, *Fundamentals of Queueing Theory*. New York: Wiley, 1974.
- [2] A. Clarke, "Maximum likelihood estimates in a simple queue," *Ann. Math. Stat.*, vol. 28, pp. 1036-1040, 1957.
- [3] U. Bhat and S. Rao, "Statistical analysis of queueing systems," *Queueing Syst.*, vol. 1, pp. 217-247, 1987.
- [4] R. Wolff, "Problems of statistical inference for birth and death queueing models," *Oper. Res.*, vol. 13, pp. 343-357, 1965.
- [5] I. Basawa and N. Prabhu, "Estimation in single server queues," *Naval Res. Log. Quart.*, vol. 28, pp. 475-487, Sept. 1981.
- [6] D. Gaver and P. Jacobs, "On inference concerning time-dependent queue performance: The M/G/1 example," *Queueing Syst.*, vol. 6, pp. 261-276, 1990.
- [7] CCITT Study Group XVIII, Rec. 1.121-Broadband Aspects of ISDN, July 1988.
- [8] T. Chen, J. Walrand, and D. Messerschmitt, "Dynamic priority protocols for packet voice," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 632-643, June 1989.
- [9] G. Barberis and D. Pazzaglia, "Analysis and optimal design of a packet voice receiver," *IEEE Trans. Commun.*, vol. COM-28, pp. 217-227, Feb. 1980.

- [10] B. Kim, "Characterization of arrival statistics of multiplexed voice packets," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 1133-1139, 1983.
- [11] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 833-846, Sept. 1986.
- [12] T. Suda *et al.*, "Performance evaluation of a packetized voice system-simulation study," *IEEE Trans. Commun.*, vol. COM-32, pp. 97-102, Jan. 1984.
- [13] W. Montgomery, "Techniques for packet voice synchronization," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 1022-1028, Dec. 1983.
- [14] D. Cox and H. Miller, *The Theory of Stochastic Processes*. New York: Wiley, 1965.
- [15] D. Cox and V. Isham, "The virtual waiting time and related processes," *Adv. Appl. Prob.*, vol. 18, pp. 558-573, 1986.
- [16] S. Rao, *Optimization: Theory and Applications*. New York: Wiley, 1984.
- [17] G. Walsh, *Methods of Optimization*. New York: Wiley, 1975.
- [18] I. Basawa and B. Prakasa Rao, *Statistical Inference for Stochastic Processes*. New York: Academic, 1980.
- [19] D. Cox and D. Hinkley, *Theoretical Statistics*. London, England: Chapman and Hall, 1974.
- [20] S. Wilks, *Mathematical Statistics*. New York: Wiley, 1962.
- [21] W. Feller, *An Introduction to Probability Theory and Its Applications*. Vol. II. New York: Wiley, 1971.
- [22] J. Kleijnen, *Statistical Tools for Simulation Practitioners*. New York: Marcel Dekker, 1987.
- [23] A. Law, "Statistical analysis of simulation output data," *Oper. Res.*, vol. 31, pp. 983-1029, Nov. 1983.
- [24] T. Chen, "Queueing and statistical issues in packet-switched voice networks," Ph.D. dissertation, Dep. EECS, Univ. of Calif., Berkeley, 1990.

Thomas M. Chen (M'88) received the M.S. and B.S. degrees from Massachusetts Institute of Technology in 1984 and the Ph.D. degree in electrical engineering from the University of California at Berkeley in 1990.

He is currently a Member of the Technical Staff at GTE Laboratories, Waltham, MA, involved in research in ATM switching systems and traffic control.

Dr. Chen is an active member of the IEEE Communications and Computer Societies.



Jean Walrand (S'71-M'74-SM'90-F'93) received the Ingénieur Civil degree from the Université de Liège (Liège, Belgium) in 1974, and the Ph.D. degree from the University of California at Berkeley in 1979, both in electrical engineering.

From 1979 to 1981 he taught at Cornell University, Ithaca, NY, and he has been since 1981 with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley where he is now Professor. His research interests are in queueing networks, communication

networks, stochastic systems, simulation, and traffic systems.

Dr. Walrand is the author of the textbooks *An Introduction to Queueing Networks* (Englewood Cliffs, NJ: Prentice Hall, 1988) and *Communication Networks: A First Course* (Irwin/Aksen, 1991). He has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL AND SYSTEMS and *Control Letters*, and is on the editorial boards of *Mathematics of Operations Research*, *Queueing Systems*, and *Probability in the Engineering and Informational Sciences*. He is a Fellow of the Belgian-American Educational Foundation. He was awarded the Lanchester Prize by the Operations Research Society of America in 1990.

David G. Messerschmitt (S'65-M'68-SM'78-F'83) for a biography and photograph, please see the January 1994 issue of this TRANSACTIONS, p.69.