Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

Estimating Function Methods Of Inference For Queueing Parameters

I. V. Basawa
The University of Georgia

Robert Lund
The University of Georgia

U. Narayan Bhat
Southern Methodist University

ABSTRACT

This paper develops estimates of the interarrival and service time distribution parameters in a GI/G/1 queueing system from observations of the waiting times of the first N customers. Specifically, if I_k and S_k denote the interarrival and service times of the kth customer arriving at the queue, then the waiting time sequence $\{W_k\}$ evolves via the Markovian recursion $W_k = \max(W_{k-1} + S_{k-1} - I_k, 0)$ for $k \geq 2$.

We first exploit the Markov structure of $\{W_k\}$ to derive an estimating function equation involving the waiting time data; in principle, this equation can be used to obtain estimates of the parameters governing the distributions of S_1 and I_1. Next, all quantities involved in the estimating function equation are expressed in terms of the distributions of S_1 and I_1. The above estimating techniques are explored in depth for the $M/E_k/1$ queue; here, explicit computations permit a simulation study of this queueing system. Finally, the consistency and asymptotic normality of the estimating function parameter estimates are established.

Key Words: Queue; waiting time; estimating function; maximum likelihood.

1 Introduction

Consider a GI/G/1 queueing system and suppose that successive customers arrive at the queue at the times $\{T_i, i \geq 1\}$. Let S_i denote the service time of the ith customer, $i \geq 1$, and define $I_i = T_i - T_{i-1}$ as the interarrival time between the ith and $(i-1)$st customers (we take $T_0 = 0$). The waiting time of the ith customer, denoted W_i, is the total amount of time the ith customer spends waiting for his/her service to commence. The waiting time process $\{W_t, t \geq 1\}$ evolves via the well-known Lindley recursion

$$W_{t+1} = \max(W_t + X_{t+1}, 0) \tag{1.1}$$

for $t \geq 1$ where $X_t = S_{t-1} - I_t$ (see Prabhu, 1980). Since X_{t+1} is independent of W_t, (1.1) shows that $\{W_t\}$ is a Markov chain on the state space $[0, \infty)$.

We will henceforth assume that the traffic intensity of the queue is subcritical; that is, $\rho = E[S_1]/E[I_1] < 1$. When $\rho < 1$, it is known that W_t converges in total variation as $t \to \infty$ to a random variable W at a geometric rate (cf. Lund, 1996). We denote the measure associated with W by π (stationary distribution) and comment that the geometric convergence and the inference procedures described below are valid for any initial distribution of W_1 satisfying $E[r^{W_1}] < \infty$ for some $r > 1$. Hence, for simplicity, we take $W_1 = 0$; that is, we assume that the queue is initially empty unless otherwise stated.

The objective of this paper is to develop estimates for the parameters of the distributions of I_1 and S_1 based on the waiting time observations W_t for $1 \leq t \leq N$. Both maximum likelihood and estimating function approaches are considered and compared; hence, this paper extends the work of Basawa *et al.* (1996).

Most previous parameter inference procedures for queueing models require the observation of all customer interarrival and service times. We refer the reader to Basawa and Prabhu (1981, 1988), Bhat and Rao (1987), Basawa and Bhat (1992), and Thiruvaiyaru and Basawa (1992) for methods and related references on this topic. Unfortunately, the observation of all interarrival and service times is frequently impractical or costly; however, the customer waiting times can easily be measured by putting a "clock" on each customer. Hence, inference procedures based only on waiting time data are often desirable and cost efficient.

In Section 2, we present the relevant theory needed to derive estimating function and maximum likelihood estimates of the interarrival and service time distribution parameters from the waiting time data. Section 3 explicitly computes the relevant quantities appearing in Section 2 equations for the case of an $M/E_k/1$ queue. Section 4 establishes the consistency and asymptotic normality of the estimating function estimates. Section 5 uses the results of Section 3 for a simulation study of the $M/E_k/1$ queue. Finally, Section 6 concludes with a summary and some comments.

We refer to Godambe and Heyde (1987), Greenwood and Wefelmeyer (1991), and Hutton and Nelson (1986) for more general treatments of estimating functions for stochastic processes.

2 General Theory

In this section, we present the general theory needed to obtain estimating function and maximum likelihood estimates of the parameters governing the interarrival and service-time distributions. For notation, let $\boldsymbol{\theta}$ denote the vector of interarrival and service time distribution parameters.

Following Godambe (1985), we define an estimating function in the form

$$S_N(\boldsymbol{\theta}) = \sum_{t=1}^{N-1} (W_{t+1} - E_{\boldsymbol{\theta}}[W_{t+1}|W_1, ..., W_t])h_t(W_1, ..., W_t; \boldsymbol{\theta}), \qquad (2.1)$$

where h_t is a function of $W_1, ..., W_t$, and $\boldsymbol{\theta}$ for $1 \leq t \leq N-1$; the subscript $\boldsymbol{\theta}$ in (2.1) indicates that the expectation is to be computed when the true parameter vector is $\boldsymbol{\theta}$. The Markov property of $\{W_t\}$ gives $E_{\boldsymbol{\theta}}[W_{t+1}|W_1, ..., W_t] = E_{\boldsymbol{\theta}}[W_{t+1}|W_t]$ and the best choice of the $h_t(\cdot)$'s are known to be (see Godambe (1985) for a definition of best)

$$h_t(W_1, ..., W_t; \boldsymbol{\theta}) = \frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}} Var_{\boldsymbol{\theta}}^{-1}[W_{t+1}|W_t]. \qquad (2.2)$$

Hence, an estimating function estimate of $\boldsymbol{\theta}$ based on $W_1, ..., W_N$ is a solution to the equation

$$S_N(\boldsymbol{\theta}) = \sum_{t=1}^{N-1} (W_{t+1} - E_{\boldsymbol{\theta}}[W_{t+1}|W_t]) \frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}} Var_{\boldsymbol{\theta}}^{-1}[W_{t+1}|W_t] = 0. \qquad (2.3)$$

A simpler estimating function estimate can also be obtained when "variance weights" in (2.3) are neglected. Hence, we will also explore solutions to the equation

$$S_N^*(\boldsymbol{\theta}) = \sum_{t=1}^{N-1} (W_{t+1} - E_{\boldsymbol{\theta}}[W_{t+1}|W_t]) \frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}} = 0. \qquad (2.4)$$

For (2.3) and (2.4) to be useful, one must be able to compute the expectations and variances appearing in these equations in terms of the interarrival and service time distributions. For this, let X be a random variable whose distribution is the same as that of $S_1 - I_1$. Define $F_X(x) = P[X \leq x]$, note that F_X depends on $\boldsymbol{\theta}$, and use (1.1) to get

$$E_{\boldsymbol{\theta}}[W_{t+1}|W_t] = W_t\alpha(W_t) + \beta(W_t), \tag{2.5}$$

where

$$\alpha(W_t) = \int_{\{x>-W_t\}} dF_X(x) \text{ and } \beta(W_t) = \int_{\{x>-W_t\}} xdF_X(x). \tag{2.6}$$

Similar arguments show that

$$E_{\boldsymbol{\theta}}[W_{t+1}^2|W_t] = W_t^2\alpha(W_t) + 2W_t\beta(W_t) + \gamma(W_t), \tag{2.7}$$

where

$$\gamma(W_t) = \int_{\{x>-W_t\}} x^2 dF_X(x). \tag{2.8}$$

From (2.7) and (2.5), we obtain

$$\begin{aligned} Var_{\boldsymbol{\theta}}[W_{t+1}|W_t] &= W_t^2\alpha(W_t)[1-\alpha(W_t)] \\ &+2W_t\beta(W_t)[1-\alpha(W_t)] + \gamma(W_t) - \beta^2(W_t). \end{aligned} \tag{2.9}$$

Notice that in principal, (2.5) - (2.9) identify all quantities in (2.3) and (2.4) in terms of the interarrival and service time distributions; in practice, one would need explicit expressions for $E_{\boldsymbol{\theta}}[W_{t+1}|W_t]$, $\frac{d}{d\boldsymbol{\theta}}E_{\boldsymbol{\theta}}[W_{t+1}|W_t]$, and $Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]$ in terms of $\boldsymbol{\theta}$ to implement (2.3) and (2.4).

Now consider the method of maximum likelihood. For simplicity, we assume that F_X has the probability density function

$$f_X(x) = \frac{d}{dx}(F_X(x)). \tag{2.10}$$

Define the indicator variable

$$Z_t = \mathcal{I}_{(0,\infty)}(W_t); \tag{2.11}$$

the Markov property of $\{W_t\}$ can be used to show that the likelihood function, denoted $L(\boldsymbol{\theta}; W_1, ..., W_N)$, satisfies

$$\begin{aligned} \log(L(\boldsymbol{\theta}; W_1, ..., W_N)) &= \sum_{t=1}^{N-1} (1 - Z_{t+1}) \log[1-\alpha(W_t)] \\ &+ \sum_{t=1}^{N-1} Z_{t+1} \log f_X(W_{t+1} - W_t) \end{aligned} \tag{2.12}$$

(see Basawa *et al.* (1996) for the details). Notice that the quantities in (2.12) are easily expressed in terms of the distribution of X.

Note that $S_N(\theta)$ in (2.3) is optimal in the class of estimating functions specified by (2.1). However, if the choice of estimating functions is not restricted to the class in (2.1), Godambe (1960) has shown, under very general conditions, that the likelihood score function, viz. $\frac{d\log L}{d\theta}$, is a (globally) optimal estimating function. In our problem, the likelihood score does not satisfy (2.1), and hence $S_N(\theta)$ is "less optimal" than $\frac{d\log L}{d\theta}$, in the sense of information content. Consequently, there is some loss of efficiency in using $S_N(\theta)$ (or $S_N^*(\theta)$) instead of $\frac{d\log L}{d\theta}$. On the otherhand, $\frac{d\log L}{d\theta}$ requires the knowledge of the density $f_X(.)$, where as $S_N(\theta)$ needs only the conditional mean and variance of $\{W_t\}$, and $S_N^*(\theta)$ requires the conditional mean only. The simulaiton results in Section 5 show that the estimates obtained from $S_N(\theta)$ and $S_N^*(\theta)$ are less biased than the maximum likelihood estimates. Moreover, the loss of efficiency due to using the estimating functions is negligible except when the traffic intensity is large.

3 Computations for the $M/E_k/1$ Queue

In this section, all quantities appearing in the Section 2 estimating function and likelihood equations (2.3), (2.4), and (2.12) will be explicitly computed in terms of $\boldsymbol{\theta}$ for the $M/E_k/1$ queue; one obtains results for the classical $M/M/1$ queue by taking $k = 1$. In the $M/E_k/1$ queue, the customer interarrival times $\{I_j\}$ are exponentially distributed random variables with parameter λ and the service times $\{S_j\}$ have the Erlang (k, μ) density. Hence, the probability density functions of I_1 and S_1, denoted by $f_I(x)$ and $f_S(x)$ respectively, are

$$f_I(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 \quad \text{and} \quad f_S(x) = \frac{\mu e^{-\mu x}(\mu x)^{k-1}}{(k-1)!} \quad \text{for } x \geq 0. \tag{3.1}$$

Straightforward computations provide the cumulative distribution function of $S_1 - I_1$:

$$\begin{aligned} F_X(x) &= \mu^k(\lambda+\mu)^{-k}e^{\lambda x}, & x < 0 \quad \text{and} \\ F_X(x) &= 1 - e^{-\mu x}\sum_{r=0}^{k-1}\frac{(\mu x)^r + \mu^k(\lambda+\mu)^{-k}[(\lambda+\mu)x]^r}{r!}, & x \geq 0. \end{aligned} \tag{3.2}$$

The probability density function $f_X(x)$ is easily obtained by differentiating (3.2):

$$\begin{aligned} f_X(x) &= \lambda\mu^k(\lambda+\mu)^{-k}e^{\lambda x}, & x < 0 \quad \text{and} \\ f_X(x) &= 1 - \lambda\mu^k(\lambda+\mu)^{-k}e^{-\mu x}\sum_{r=0}^{k-1}\frac{[(\lambda+\mu)x]^r}{r!}, & x \geq 0. \end{aligned} \tag{3.3}$$

From the first expression in (3.2), we obtain

$$\alpha(W_t) = 1 - (\frac{\mu}{\lambda+\mu})^k e^{-\lambda W_t}. \tag{3.4}$$

More tedious computations with the density function in (3.3) give

$$\beta(W_t) = \lambda^{-1}(\frac{\mu}{\lambda+\mu})^k e^{-\lambda W_t}(1+\lambda W_t) + \mu^{-1}k - \lambda^{-1} \tag{3.5}$$

and

$$\gamma(W_t) = \lambda^{-2}(\frac{\mu}{\lambda+\mu})^k e^{-\lambda W_t}[-\lambda^2 W_t^2 + 2\lambda W_t - 2]$$
$$+\mu^{-2}k(k+1) - 2\lambda^{-2}(k\lambda/\mu - 1). \tag{3.6}$$

From (2.5) and (2.9), we see that $E_{\boldsymbol{\theta}}[W_{t+1}|W_t]$ and $Var_{\boldsymbol{\theta}}(W_{t+1}|W_t)$ are easily obtained in terms of $\alpha(W_t)$, $\beta(W_t)$, and $\gamma(W_t)$. To complete the computation of all quantities in (2.3), (2.4), and (2.12), we must evaluate the partial derivatives of $E_{\boldsymbol{\theta}}[W_{t+1}|W_t]$ with respect to λ and μ. Using (2.5), (3.4), (3.5), and the notation $E_{\boldsymbol{\theta}}[W_{t+1}|W_t] = E_{\lambda,\mu}[W_{t+1}|W_t]$, we find that

$$\frac{d}{d\lambda}E_{\lambda,\mu}[W_{t+1}|W_t] = \lambda^{-2} - \lambda^{-2}(\frac{\mu}{\lambda+\mu})^k e^{-\lambda W_t}[\lambda W_t + 1 + (\lambda+\mu)^{-1}\lambda k]; \tag{3.7}$$

$$\frac{d}{d\mu}E_{\lambda,\mu}[W_{t+1}|W_t] = k\mu^{-2}\left[(\frac{\mu}{\lambda+\mu})^{k+1}e^{-\lambda W_t} - 1\right].$$

4 Asymptotic Properties of the Estimates

We now follow Klimko and Nelson (1978), Hutton and Nelson (1986), and Hutton et al. (1991) and establish the consistency and asymptotic normality of the estimates obtained as solutions to (2.3) and (2.4). We will focus on the asymptotic properties of the estimating function estimates only and refer the reader to Basawa *et al.* (1996) for the asymptotic properties of the maximum likelihood estimates.

From (2.3) and (2.4), it is straightforward to show that $\{S_N(\boldsymbol{\theta})\}$ and $\{S_N^*(\boldsymbol{\theta})\}$ are mean zero martingales with respect to $\{W_t\}$. Two results, Lemmas 4.1 and 4.2 below, that will be helpful later are now stated. The expectations in (4.1)-(4.4) are tacitly assumed to exist and are taken with respect to the stationary measure π.

LEMMA 4.1. Consider the waiting time process $\{W_t\}$ in (1.1) with $\rho < 1$. Let $S_N(\boldsymbol{\theta})$ and $S_N^*(\boldsymbol{\theta})$ be defined as in (2.3) and (2.4); then the following convergence takes place in probability as $N \to \infty$.

(*i*) $N^{-1}S_N(\boldsymbol{\theta}) \xrightarrow{P} 0$ and $N^{-1}S_N^*(\boldsymbol{\theta}) \xrightarrow{P} 0$.

(*ii*) $-N^{-1}\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \xrightarrow{P} J(\boldsymbol{\theta})$ and $-N^{-1}\frac{dS_N^*(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \xrightarrow{P} J^*(\boldsymbol{\theta})$ where

$$J(\boldsymbol{\theta}) = E\left[\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)\left(\frac{\frac{d}{d\boldsymbol{\theta}}E_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]}\right)'\right], \text{ and} \tag{4.1}$$

$$J^*(\boldsymbol{\theta}) = E\left[\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)'\right]. \tag{4.2}$$

PROOF: Since $\{W_t\}$ is an ergodic process when $\rho < 1$, the results in (i) and (ii) will follow from the ergodic theorem and algebraic computations with (2.3) and (2.4). We shall verify the results for $S_N(\boldsymbol{\theta})$. Similar arguments can be used for $S_N^*(\boldsymbol{\theta})$.

Define

$$U_t(\boldsymbol{\theta}) = (W_{t+1} - E_{\boldsymbol{\theta}}[W_{t+1}|W_t])\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}Var_{\boldsymbol{\theta}}^{-1}[W_{t+1}|W_t],$$

we have $S_N(\boldsymbol{\theta}) = \sum_{t=1}^{N-1} U_t(\boldsymbol{\theta})$. Clearly, $EU_t(\boldsymbol{\theta}) = 0$. The ergodic theorem then gives the result in (i). Also,

$$\begin{aligned}\frac{dU_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}} &= - (\frac{d}{d\boldsymbol{\theta}}E_{\boldsymbol{\theta}}[W_{t+1}|W_t])\left(\frac{\frac{d}{d\boldsymbol{\theta}}E_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]}\right)' \\ &+ (W_{t+1} - E_{\boldsymbol{\theta}}[W_{t+1}|W_t])(\frac{d}{d\boldsymbol{\theta}}\{\frac{\frac{d}{d\boldsymbol{\theta}}E_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]}\}).\end{aligned}$$

It then follows readily that

$$E_\theta(-\frac{dU_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}}) = J(\boldsymbol{\theta}).$$

Hence, (ii) follows from the ergodic theorem. □

LEMMA 4.2. Under the notation and assumptions of Lemma 5.1, the following convergence takes place in distribution as $N \to \infty$.
(i) $N^{-1/2}S_N(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(0, F(\boldsymbol{\theta}))$ where

$$F(\boldsymbol{\theta}) = E\left[Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]^{-1}\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)'\right]. \tag{4.3}$$

(ii) $N^{-1/2}S_N^*(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(0, F^*(\boldsymbol{\theta}))$ where

$$F^*(\boldsymbol{\theta}) = E\left[Var_{\boldsymbol{\theta}}[W_{t+1}|W_t]\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)\left(\frac{dE_{\boldsymbol{\theta}}[W_{t+1}|W_t]}{d\boldsymbol{\theta}}\right)'\right]. \tag{4.4}$$

PROOF: Notice that $S_N(\boldsymbol{\theta})$ and $S_N^*(\boldsymbol{\theta})$ are sums of stationary ergodic martingale differences with finite second moments. An appeal to the martingale central limit theorem (cf. Billingsley (1961)) easily establishes (i) and (ii). Note that

$$E_\theta(U_t(\boldsymbol{\theta})U_t'(\boldsymbol{\theta})) = F(\boldsymbol{\theta}).$$

A similar computation holds for $F^*(\boldsymbol{\theta})$. □

Note that $F(\boldsymbol{\theta}) = J(\boldsymbol{\theta})$, however, $F^*(\boldsymbol{\theta}) \neq J^*(\boldsymbol{\theta})$. We will now confine our attention to $S_N(\boldsymbol{\theta})$; analogous results for $S_N^*(\boldsymbol{\theta})$ can be obtained by "starring" all quantities in the results below. Consider the following two conditions.

(C.1) Suppose $S_N(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, and for all $\delta > 0$,

$$P\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|=\delta} (\boldsymbol{\theta}-\boldsymbol{\theta}_0)' S_N(\boldsymbol{\theta}) < -\epsilon\right) \to 1, \tag{4.5}$$

for any $\epsilon > 0$.

(C.2) Suppose that if $\{\tilde{\boldsymbol{\theta}}_N\}$ is any sequence of estimates such that $\tilde{\boldsymbol{\theta}}_N \xrightarrow{P} \theta_0$, then

$$N^{-1}\left[\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_N} - \frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \xrightarrow{P} 0 \text{ as } N \to \infty. \tag{4.6}$$

Condition (C.2) imposes a type of continuous convergence on $\{\frac{d}{d\boldsymbol{\theta}} S_N(\boldsymbol{\theta})\}$. Sufficient conditions for (C.2) to hold can be phrased in terms of expectations of the second derivative of $S_N(\boldsymbol{\theta})$. The interested reader is referred to Klimko and Nelson (1978) for further details. See Hutton et al. (1991) for sufficient conditions for (C.1). We note that (C.1) and (C.2) can be verified for the $M/M/1$ and $M/E_k/1$ queues when $\rho < 1$ from these second derivative conditions and the equations in Sections 2 and 3.

Our next two results establish the consistency and asymptotic normality of the estimating function estimates.

THEOREM 4.1. Let $\{W_t\}$ be the waiting time process in (1.1) with $\rho < 1$, and suppose that $S_N(\boldsymbol{\theta})$ in (2.3) satisfies (C.1). Then there exists a sequence of estimators $\hat{\boldsymbol{\theta}}_N$ such that $P_{\boldsymbol{\theta}_0}[S_N(\hat{\boldsymbol{\theta}}_N) = 0] \to 1$ as $N \to \infty$ and $\hat{\boldsymbol{\theta}}_N \xrightarrow{P} \boldsymbol{\theta}_0$ as $N \to \infty$.

PROOF: See Hutton et al. (1991), or Hutton and Nelson (1986).

THEOREM 4.2. If $\rho < 1$, and (C.1) and (C.2) are satisfied, and if $\hat{\boldsymbol{\theta}}_N$ is any consistent solution of $S_N(\boldsymbol{\theta}) = 0$, then

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(0, F^{-1}(\boldsymbol{\theta}_0)).$$

PROOF: A Taylor expansion of $S_N(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ gives

$$S_N(\boldsymbol{\theta}) = S_N(\boldsymbol{\theta}_0) + \left[\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right](\boldsymbol{\theta}-\boldsymbol{\theta}_0) \tag{4.7}$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Replacing $\boldsymbol{\theta}$ in (4.7) by $\hat{\boldsymbol{\theta}}_N$, we have

$$0 = S_N(\boldsymbol{\theta}_0) + \left[\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_N}\right](\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0), \tag{4.8}$$

where $\tilde{\boldsymbol{\theta}}_N$ lies between $\hat{\boldsymbol{\theta}}_N$ and $\boldsymbol{\theta}_0$. From (4.8), we obtain

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) = -\left[N^{-1}\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_N}\right]^{-1} N^{-1/2}S_N(\boldsymbol{\theta}_0). \qquad (4.9)$$

Lemma 4.2 shows that

$$N^{-1/2}S_N(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(0, F(\boldsymbol{\theta}_0)). \qquad (4.10)$$

From Lemma 4.1 *(ii)* and (C.2), we have

$$-N^{-1}\frac{dS_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_N} \xrightarrow{\mathcal{D}} J(\boldsymbol{\theta}_0) = F(\boldsymbol{\theta}_0). \qquad (4.11)$$

Combining (4.9) - (4.11), we obtain the desired result. □

Following similar arguments, we have

THEOREM 4.3. If (C.1) and (C.2) are satisfied and if $\hat{\boldsymbol{\theta}}^*_N$ is any consistent solution of $S^*_N(\boldsymbol{\theta}) = 0$, then

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^*_N - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(0, (J^{*\prime}(\boldsymbol{\theta}_0)F^*(\boldsymbol{\theta}_0)^{-1}J^*(\boldsymbol{\theta}_0)^{-1}).$$

We comment that it would be a straightforward, albeit tedious, matter to derive explicit expressions for $J(\boldsymbol{\theta})$ and $F(\boldsymbol{\theta})$ for the $M/E_k/1$ queue. For example, in the $M/M/1$ queue, one could use that the limiting measure π has an atom at $\{0\}$ and is exponentially distributed elsewhere (cf. Prabhu (1980)); specifically, $\pi(\{0\}) = 1-\lambda\mu^{-1}$ and $\pi(dx) = \lambda\mu^{-1}e^{-(\mu-\lambda)x}$ for $x > 0$. This could be combined with the equations in Sections 2 and 3 to obtain $E_{\lambda,\mu}[W_{t+1}|W_t]$ when W_t has distribution π. Similar computations would give $Var_{\lambda,\mu}[W_{t+1}|W_t]$ and (4.1)-(4.4) could then be used to compute $J(\boldsymbol{\theta})$ and $F(\boldsymbol{\theta})$. These details are omitted.

5 A Simulation Study

In this section, we will compare properties of the estimating function estimates (with and without variance weights) and the method of maximum likelihood estimates via simulation. Waiting time data were simulated and the performance of the estimating methods was investigated. The results are summarized in Tables 1, 2, and 3 below.

Table 1 considers the M/M/1 queue. The parameter pairs $\lambda = 1$, $\mu = 2$; $\lambda = 2$, $\mu = 3$; and $\lambda = 5$, $\mu = 6$ were studied; these parameter pairs yield the increasing traffic intensities $\rho = 1/2, 2/3$, and $5/6$ respectively. One hundred simulations were performed for each (λ, μ) pair and each of the sample sizes $N = 100$, $N = 250$, and $N = 500$. Table 1 shows the sample mean and root mean squared errors for each simulation. A separate subtable is included for each of the three methods of estimation.

TABLE 1: The M/M/1 queue.

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ). Method: estimating function with variance weights.

	Sample Size N			
	100	250	500	1000
$\lambda = 1$	0.915 (0.414)	1.007 (0.313)	0.961 (0.209)	1.030 (0.158)
$\mu = 2$	1.917 (0.606)	2.023 (0.426)	1.957 (0.266)	2.044 (0.221)
$\lambda = 2$	1.780 (0.859)	1.914 (0.543)	1.969 (0.355)	2.016 (0.243)
$\mu = 3$	2.863 (0.959)	2.901 (0.604)	3.024 (0.416)	3.047 (0.274)
$\lambda = 5$	5.034 (3.072)	5.050 (1.767)	4.902 (0.974)	4.992 (0.663)
$\mu = 6$	5.974 (2.557)	6.095 (1.673)	5.949 (0.985)	5.975 (0.760)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ). Method: maximum likelihood.

	Sample Size N			
	100	250	500	1000
$\lambda = 1$	1.439 (0.533)	1.423 (0.464)	1.401 (0.416)	1.407 (0.417)
$\mu = 2$	2.311 (0.466)	2.309 (0.388)	2.285 (0.341)	2.269 (0.289)
$\lambda = 2$	2.671 (0.816)	2.609 (0.669)	2.625 (0.648)	2.550 (0.563)
$\mu = 3$	3.488 (0.751)	3.368 (0.497)	3.330 (0.414)	3.285 (0.331)
$\lambda = 5$	6.035 (1.302)	5.790 (0.922)	5.706 (0.765)	5.711 (0.739)
$\mu = 6$	6.435 (1.077)	6.386 (0.674)	6.309 (0.497)	6.290 (0.393)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: estimating function without variance weights.

	Sample Size N			
	100	250	500	1000
$\lambda = 1$	0.978 (0.413)	1.032 (0.289)	0.999 (0.203)	0.970 (0.145)
$\mu = 2$	1.947 (0.522)	2.079 (0.382)	2.017 (0.274)	1.944 (0.206)
$\lambda = 2$	1.858 (0.729)	1.889 (0.474)	1.994 (0.363)	1.992 (0.292)
$\mu = 3$	2.850 (0.805)	2.943 (0.539)	3.039 (0.409)	3.004 (0.333)
$\lambda = 5$	4.747 (2.286)	5.036 (1.811)	4.737 (0.926)	5.072 (0.775)
$\mu = 6$	5.733 (2.255)	6.110 (1.681)	5.742 (0.938)	6.064 (0.776)

Table 1 shows that the two estimating functions methods yield approximately unbiased parameter estimates; in contrast, all maximum likelihood sample means are larger than the true parameter values. Despite this bias, the method of maximum likelihood has a smaller root mean squared error than both estimating function methods for the traffic intensity $\rho = 5/6$. This, of course, reflects the fact that the maximum likelihood estimates had a much smaller sample variance than their estimating function counterparts. We note that for $\rho = 1/2$, the root mean squared errors from the estimating function methods are comparable (sometimes even smaller) to the maximum likelihood estimate root mean squared errors. Inspection of Table 1 shows that the estimating function estimates without variance weights are, overall, about as efficient as the estimating function estimates with variance weights in terms of root mean squared error. Finally, we note that the root mean squared errors of all estimates increase with increasing λ and/or μ.

In terms of computations, the maximum likelihood estimates were the easiest to obtain. The minimum of the negative log likelihood function was rapidly found in all simulations with a gradient search routinge. In contrast, difficulties were encountered with the root finding computations needed to compute the estimating function estimates. In a small proportion of the simulations with the smaller series lengths (particularly $N = 100$), none of the standard root finding numerical methods tried such as Newton or Broyden satisfactorily found the roots in all simulations. The root finding method that worked best in practice proceeded as follows. First, a gradient search routing was used to find "approximate roots" of the estimating function equations by numerically minimizing the sum of squares

$$g(\lambda, \mu) = S_N^{(1)}(\lambda, \mu)^2 + S_N^{(2)}(\lambda, \mu)^2$$

in λ and μ, where $S_N^{(i)}(\lambda,\mu)$ is the ith component of $S_N(\boldsymbol{\theta})$ or $S_N^*(\boldsymbol{\theta})$ for $i = 1, 2$ ($i = 1$ corresponds to λ, $i = 2$ corresponds to μ). These estimates were then refined with Newton's method for systems of non-linear equations. In virtually all simulations, a root $(\hat{\lambda}, \hat{\mu})$ satisfying the tolerance

$$|S_N^{(1)}(\hat{\lambda},\hat{\mu})| + |S_N^{(2)}(\hat{\lambda},\hat{\mu})| \leq 10^{-7}$$

was found.

Tables 2 and 3 show similar simulations for the $M/E_k/1$ queue for the cases $k = 2$ and $k = 4$ respectively. The parameter values of λ and μ were again selected to yield the increasing traffic intensities $\rho = 1/2, 2/3$, and $5/6$. The bias properties of the estimates in Tables 2 and 3 are similar to those in Table 1. We note that the maximum likelihood estimates, in most cases, have smaller root mean squared errors than their estimating function counterparts. In many cases, the estimating function approach without variance weights yielded a root mean squared error that was comparable, or only slightly larger, to the root mean squared error of the estimating function approach without variance weights. Hence, little seems to be gained by accounting for variances in the estimating function approach for the $M/E_k/1$ queue. It should be noted, however, that the gain in efficiency due to accounting for variance weights should increase with larger sample sizes.

TABLE 2: The $M/E_2/1$ queue.

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: estimating function with variance weights.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.005 (0.395)	0.971 (0.283)	0.965 (0.230)	0.963 (0.150)
$\mu = 4$	4.045 (1.140)	3.990 (0.681)	3.947 (0.599)	3.961 (0.407)
$\lambda = 2$	1.913 (0.912)	1.976 (0.541)	1.935 (0.347)	1.986 (0.230)
$\mu = 6$	5.917 (1.974)	5.967 (1.163)	5.888 (0.768)	5.968 (0.496)
$\lambda = 5$	4.827 (2.059)	4.925 (1.279)	5.044 (0.990)	4.932 (0.598)
$\mu = 12$	11.975 (4.062)	11.947 (2.484)	12.167 (2.108)	11.922 (1.220)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: maximum likelihood.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.386 (0.457)	1.411 (0.450)	1.393 (0.410)	1.337 (0.375)
$\mu = 4$	4.447 (0.731)	4.513 (0.661)	4.516 (0.602)	4.504 (0.548)
$\lambda = 2$	2.675 (0.775)	2.633 (0.667)	2.596 (0.622)	2.579 (0.592)
$\mu = 6$	6.808 (1.205)	6.640 (0.841)	6.628 (0.738)	6.552 (0.616)
$\lambda = 5$	6.046 (1.281)	5.684 (0.981)	5.779 (0.833)	5.760 (0.791)
$\mu = 12$	12.915 (1.677)	12.917 (1.398)	12.643 (0.927)	12.616 (0.783)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: estimating function without variance weights.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.050 (0.489)	0.955 (0.299)	0.957 (0.202)	0.984 (0.165)
$\mu = 4$	4.129 (1.366)	3.914 (0.824)	3.894 (0.560)	3.964 (0.449)
$\lambda = 2$	2.023 (0.836)	1.998 (0.468)	1.994 (0.352)	1.981 (0.255)
$\mu = 6$	6.119 (1.797)	6.036 (1.095)	5.989 (0.787)	5.935 (0.528)
$\lambda = 5$	5.249 (2.314)	5.176 (1.324)	5.051 (0.888)	5.030 (0.590)
$\mu = 12$	12.733 (4.046)	12.419 (2.690)	12.016 (1.743)	12.066 (1.193)

TABLE 3: The $M/E_4/1$ queue.

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: estimating function with variance weights.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.099 (0.495)	0.958 (0.286)	0.969 (0.222)	0.990 (0.166)
$\mu = 8$	8.650 (2.759)	7.889 (1.453)	7.869 (1.143)	7.971 (0.828)
$\lambda = 2$	1.811 (0.748)	1.910 (0.481)	1.931 (0.329)	1.946 (0.260)
$\mu = 12$	11.636 (3.522)	11.701 (1.929)	11.775 (1.323)	11.831 (1.123)
$\lambda = 5$	4.788 (2.316)	4.842 (1.262)	5.091 (0.948)	4.921 (0.592)
$\mu = 24$	23.549 (7.694)	23.609 (5.074)	24.334 (3.475)	23.850 (2.240)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: maximum likelihood.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.362 (0.438)	1.324 (0.356)	1.338 (0.355)	1.326 (0.334)
$\mu = 8$	8.903 (1.255)	8.758 (0.996)	8.840 (0.950)	8.795 (0.849)
$\lambda = 2$	2.540 (0.645)	2.588 (0.633)	2.544 (0.570)	2.538 (0.549)
$\mu = 12$	13.100 (1.673)	13.049 (1.360)	12.969 (1.143)	12.907 (0.974)
$\lambda = 5$	5.926 (1.106)	5.800 (0.895)	5.782 (0.825)	5.785 (0.808)
$\mu = 24$	25.412 (2.811)	25.261 (2.012)	25.290 (1.697)	25.145 (1.393)

Sample mean and root mean squared error of 100 simulated parameter estimates of (λ, μ).
Method: estimating function without variance weights.

Sample Size N

	100	250	500	1000
$\lambda = 1$	1.085 (0.543)	1.024 (0.323)	1.012 (0.350)	1.004 (0.187)
$\mu = 8$	8.241 (2.843)	8.161 (1.766)	7.982 (1.664)	8.036 (1.033)
$\lambda = 2$	2.026 (0.949)	1.959 (0.467)	1.972 (0.379)	1.961 (0.243)
$\mu = 12$	12.104 (4.014)	11.837 (2.057)	11.944 (1.662)	11.785 (1.062)
$\lambda = 5$	5.065 (1.830)	5.006 (1.418)	4.935 (0.828)	4.841 (0.666)
$\mu = 24$	23.967 (6.196)	24.103 (3.923)	23.975 (3.101)	23.466 (2.633)

6 Summary and Comments

This paper shows how parameter estimates for the interarrival and service time distributions in a GI/G/1 queue can be obtained from customer waiting time data. Both estimating function and maximum likelihood methods of estimation were considered. The simulation study in Section 4 shows that the maximum likelihood estimates can be significantly biased, while the estimating function estimates are approximately unbiased. Despite this bias, the maximum likelihood estimates had a smaller root mean squared error than their estimating function counterparts; a similar ordering of root mean squared errors did not hold for moderate traffic intensities. For the sample sizes considered in the simulation, accounting for "variances" in the estimating function produced little gain in estimation efficiency; however, it is expected that the efficiency of the variance weighted estimates would be superior with larger sample sizes. The consistency and asymptotic normality of the estimating function estimates were also established.

Acknowledgements

I. V. Basawa's work was partially supported by grants from the Office of Naval Research and the National Science Foundation. Robert Lund's research was supported by National Science Foundation Grant DMS-9703838. We thank the referee for a careful reading and some constructive suggestions.

References

Basawa, I. V. and B. R. Bhat (1992). Sequential inference for single server queues. In *Queueing and Related Models*, 325-336, Edited by U. N. Bhat and I. V. Basawa, Oxford University Press, Oxford.

Basawa, I. V., Bhat, U. N., and R. B. Lund (1996). Maximum likelihood estimation for single server queues from waiting time data, to appear in *Queueing Systems.*

Basawa, I. V. and N. U. Prabhu (1981). Estimation in single server queues, *Naval Research Logistics Quarterly*, 28, 475-487.

Basawa, I. V. and N. U. Prabhu (1988). Large sample inference from single server queues, *Queueing Systems*, **3**, 289-306.

Bhat, U. N. and S. S. Rao (1987). Statistical analysis of queueing systems, *Queueing Systems*, **1**, 217-247.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales, *Proceedings of the American Mathematical Society*, **12**, 788-792.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.*, **31**, 1208-1212.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419-428.

Godambe, V. P. and C. C. Heyde (1987). Quasilikelihood and optimal estimation. *Int. Statist. Rev.*, **55**, 231-244.

Greenwood, P. E. and W. Wefelmeyer (1991). On optimal estimating functions for partially specified counting process models. In *Estimating Functions*, Ed. V. P. Godambe, p 147-160, Oxford Univ. Press, Oxford.

Hutton, J. E. and P. I. Nelson (1986). Quasilikelihood estimation for semimartingales. *Stoch. Proc. and Applns.* 22, 245-257.

Hutton, J. E., O. T. Ogunyemi and P. I. Nelson (1991). Simplified and two-stage quasi-likelihood estimators. In *Estimating Functions*, Ed. V. P. Godambe, p 169-187, Oxford Univ. Press, Oxford.

Klimko, L. and P. I. Nelson (1978). On conditional least squares estimation for stochastic processes, *Annals of Mathematical Statistics*, **6**, 629-642.

Lund, R. B. (1996). The geometric convergence rate of a Lindley random walk, to appear, *Journal of Applied Probability.*

Prabhu, N. U. (1980). *Stochastic Storage Processes*, Springer-Verlag, New York.

Thiruvaiyaru, D. and I. V. Basawa (1992). Empirical Bayes estimation for queueing systems and networks, *Queueing Systems*, **11**, 179-202.