Constructing estimating equations for queue length data ⋆

R. McVinish *, P.K. Pollett

School of Mathematics and Physics, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia

Abstract

In this paper we consider the problem of estimating the parameters of a Markov queuing model from discrete time observations of the queue length. The proposed approach is an application of the martingale estimating function methodology which has been used extensively in mathematical finance. A small simulation study suggests that the estimator performs well, even for moderate sample size, and that it is an improvement over the Gaussian diffusion based, approximate maximum likelihood estimator.

Key words: $M/M/c$ queue, quasi-likelihood, preemptive priority queue, tandem queue

1 Introduction

Much of the literature on parameter estimation for queuing processes concerns continuous observation of the entire queue (for example Ausín et al., 2004). However, there has been some recent interest in parameter estimation based on other types of data. The problem of parameter estimation from inter-departure times for $M/G/1$ queues has been studied using a variety of techniques. Fearnhead (2004) showed that the corresponding likelihood function can be computed exactly using a filtering algorithm and hence was able

⋆ Research supported by the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.

* Corresponding author.

Email addresses: r.mcvinish@uq.edu.au (R. McVinish), pkp@maths.uq.edu.au (P.K. Pollett).

to compute the maximum likelihood estimate. Two simulation based inferential procedures have also been proposed for this problem: indirect inference (Heggland and Frigessi, 2004) and approximate Bayesian computation (Blum and Fançois, 2009). Basawa et al. (2008) demonstrated how to compute the maximum likelihood estimate for a $GI/G/1$ queue based on waiting time data, adjusted for any idle times, using the EM algorithm. Novak and Watson (2009) devised a method of moments estimator for an $M/D/1$ queue based on delay measurements from single packet probing.

Of particular interest in this paper is the method proposed by Ross et al. (2007) for estimation of the parameters of an $M/M/c$ queue from discrete time observations of the queue length. The main difficulty in parameter estimation in this setting is that, for most queuing systems, the transition probabilities for the queue length process are not easily calculated. As a result, likelihood based estimation is far from routine. Given this difficulty, Ross et al. (2007) proposed an approximate maximum likelihood estimator for $M/M/c$ queues. Based on the results of Kurtz (1971) and Barbour (1974), they proposed using the likelihood function of a Gaussian Ornstein-Uhlenbeck (OU) process as an approximation of the true likelihood function, hereafter referred to as the OU likelihood. The OU likelihood function is easily evaluated and the estimator can be obtained using standard numerical optimisation procedures. Although the estimator is easy to compute, this approach has two significant drawbacks which have been noted by the authors. Firstly, the method produces accurate estimates only when the number of servers is large ($c \gg 40$). Secondly, the proposed confidence sets appear to have rather poor coverage properties (see Figure 1 of Ross et al., 2007).

Viewing the approximate OU likelihood approach as an estimating equation (Heyde, 1997), the poor performance for moderate and small values of c is perhaps not surprising. This is due to the fact that the resulting approximate score function, the derivative of the log likelihood, will not, in general, have expectation zero. Therefore, the estimator cannot be consistent.

This paper has two aims. Firstly, an estimator for the $M/M/c$ queue is proposed that improves on the approximate OU likelihood method of Ross et al. (2007). The estimator is constructed applying the theory of martingale estimating equations. This approach combines the ideas from the OU approximation of the queue length process with simulation to ensure that the estimating equation is unbiased. The improvement over the OU likelihood approximation is demonstrated by a direct comparison on the simulated queues used in Ross et al. (2007). The second aim of the paper is to demonstrate how the methodology can be extended to deal with more complex queues. This is done for a queue with two priority classes and for a tandem queue. Other generalisations are possible such as queuing networks and bulk-arrival bulk service queues. Although the restriction to Markovian queues excludes some basic queues from consideration, such as the $GI/G/1$ queue, we believe the class of Markovian queues is sufficiently rich to be of interest. In the discussion, we provide some comments on the application of estimating equations to non-Markovian queues.

2 An overview of estimating functions

We assume that the queue length process $X(t)$ is described by a pure jump Markov process taking values in $E \subset \mathbb{Z}^d$ with transition rates $q(i,j)$. For simplicity of notation, suppose that the queuing process $X(t)$ is observed at times $1, 2, \dots$ and denote these observations by $X_0, X_1, X_2, \dots$ The queuing process is assumed to be parameterised by $\theta \in \Theta \subset \mathbb{R}^p$.

2.1 Formulating estimating equations

The following is a brief summary of the theory of estimating equations relevant to the problem at hand. The reader should consult Heyde (1997) for a more complete survey of the theory. An estimating function is a p-dimensional function of the data and parameter

$$G_T(\theta) := G_T(\theta, X_1, \dots, X_T).$$

The estimator θ^* of θ is obtained by solving the estimating equation $G_T(\theta) = 0$. This general framework incorporates most estimators including method of moments and regular maximum likelihood estimators. The estimating function in the later case, being the derivative of the log likelihood, is also known as the score function.

We focus on a special class of estimating equations called martingale estimating equations (MEE). Let $\mathcal{F}_t$ be the sigma field generated by observations $X_1, \dots, X_t$. A martingale estimating function is an estimating function that has the martingale property,

$$\mathbb{E}_\theta \left[G_T(\theta) | \mathcal{F}_{T-1} \right] = G_{T-1}(\theta).$$

Let $\{m_t(\theta)\}_{t=1}^T$ be a martingale difference sequence, that is, for each t, $m_t(\theta)$ is $\mathcal{F}_t$ measurable and satisfies

$$\mathbb{E}_\theta \left(m_t(\theta) \mid \mathcal{F}_{t-1} \right) = 0.$$

The simplest way to construct a martingale difference sequence from $\{X_t\}_{t=1}^T$ is to transform the data using some function $h(\cdot)$ and then correct for the conditional expectation. For example, if we take $h(x) = x$ then the martingale difference sequence is $m_t(\theta) := X_t - M(X_{t-1};\theta)$, where

$$M(x;\theta) = \mathbb{E}_\theta[X_t|X_{t-1} = x].$$

Consider the class of martingale estimating functions which are linear in $m_t(\theta)$. More precisely, consider the class of martingale estimating functions

$$\mathcal{G} = \left\{ G_T(\theta) : G_T(\theta) = \sum_{t=1}^{T} a_t(\theta) m_t(\theta) \right\},$$

where the $a_t(\theta)$ are $\mathcal{F}_{t-1}$ measurable $p \times m$ matrices. Naturally, we would like to chose the estimating function in $\mathcal{G}$ which is optimal in some sense. According to the fixed sample optimality criteria (Heyde, 1997, pg 11), an estimating function is optimal within the class $\mathcal{G}$ if it minimises the distance to the score function. The optimal estimating function is obtained by setting $a_t(\theta) := a_t^*(\theta)$, where

$$a_t^*(\theta) = \left(\mathbb{E}_\theta \partial m_t(\theta)|\mathcal{F}_{t-1}\right)' \left(\mathbb{E}_\theta m_t(\theta) m_t(\theta)'|\mathcal{F}_{t-1}\right)^{-1}, \tag{1}$$

and $\partial m_t(\theta)$ denotes differentiation of $m_t(\theta)$ with respect to θ. In the case where $m_t(\theta) = X_t - M(X_{t-1};\theta)$, the corresponding optimal estimating function is

$$G_T^*(\theta) = \sum_{t=1}^{T} \partial M(X_{t-1};\theta) V^{-1}(X_{t-1};\theta) \left[X_t - M(X_{t-1};\theta)\right],$$

where $\partial M(x;\theta)$ denotes the vector of partial derivatives of $M(x;\theta)$ with respect to θ and

$$V(x;\theta) = \mathbb{E}_\theta[(X_t - M(x;\theta))(X_t - M(x;\theta))'|X_{t-1} = x].$$

We note that the choice of $\mathcal{G}$ affects the efficiency of the estimator. Ideally, the score function will be an element of $\mathcal{G}$.

2.2 Inference

Estimates obtained by solving a MEE typically have nice large sample properties. In this subsection we give an informal summary of the theory underlying these properties. The verification of sufficient conditions for the asymptotic properties to hold remains a challenge and needs to be investigated a on case by case basis.

We first define the quadratic characteristic of the martingale $G_T(\theta)$ as the random positive semi-definite $p \times p$ matrix

$$\langle G(\theta)\rangle_T = \sum_{t=1}^{T} a_t(\theta)\mathbb{E}_\theta\left(m_t(\theta)m_t(\theta)'|\mathcal{F}_{t-1}\right)a_t(\theta)'.$$

Let θ_0 denote the true parameter value. Assuming that $X(t)$ is ergodic and certain moments of $X(t)$ are finite, the conditions of Theorem 2.3 of Bibby et al. (2004) hold and $G_T(\theta_0)$ converges in distribution to a normal random variable as $T \to \infty$,

$$\langle G(\theta_0)\rangle_T^{-1/2} G_T(\theta_0) \xrightarrow{d} N(0, I). \tag{2}$$

It follows from (2) that estimating functions in $\mathcal{G}$ satisfy

$$T^{-1}G_T(\theta) \xrightarrow{P_\theta} 0. \tag{3}$$

This property is desirable as it is necessary for consistency of the estimator obtained by solving $G_T(\theta) = 0$.

Assume that $X(t)$ satisfies additional regularity conditions concerning the smoothness with respect to θ of $a_t(\theta)$ and of certain conditional moments. Then Theorem 2.2 of Bibby et al. (2004) can be applied to show that the estimator is consistent, asymptotically normal. Although the asymptotic normality of the estimator can be used to construct confidence sets for the true parameter, empirical evidence favours constructing confidence sets directly from the normalised estimating function. If the convergence in (2) can be strengthened to stable convergence, then an asymptotic $(1-\alpha)$ confidence set for θ is given by

$$\left\{\theta : G_T'(\theta)\left\langle G_T(\theta)\right\rangle^{-1} G_T(\theta) \leq \chi^2_{p;(1-\alpha)}\right\}, \tag{4}$$

where $\chi^2_{p;(1-\alpha)}$ is the $1-\alpha$ percentile of the χ^2 distribution with p degrees of freedom.

2.3 Numerical solution

The optimal MEE involves quantities that are not available in closed form for most queuing systems. Therefore, the difficulties in computing the optimal MEE are similar to those in computing the likelihood function.

In the following section we derive some approximations to (1) that are easily evaluated. Replacing (1) by an approximation entails some loss of efficiency however the resulting estimating equation remains unbiased and hence the estimator should retain the nice asymptotic properties. The approximation to (1) does not completely resolve the difficulty in evaluating the MEE since the martingale difference sequence still requires the evaluation of a conditional moment. We do not replace this conditional moment with an approximation since the estimating equation would no longer satisfy (2) and the resulting estimator would not be consistent. Accurate evaluation of a conditional moment can be achieved by simulating a large number of sample paths and taking the appropriate average. This approach was used by Bibby and Sørensen (1995) to evaluate the estimating functions for diffusion processes.

An alternate approach is to use a stochastic root finding algorithm as described in Spall (2003, chapter 4) to solve the estimating equation. Assume that the martingale difference sequence that we are using is $m_t(\theta) = X_t - M(X_{t-1};\theta)$. Let $\{\tilde{X}_t\}$ be a sequence of independent random variables such that $\tilde{X}_t$ is simulated from the conditional distribution of X_t given X_{t-1}. Define

$$\tilde{G}_T(\theta) = \sum_{t=1}^{T} a_t(\theta) \left(X_t - \tilde{X}_t \right).$$

Then

$$\mathbb{E}_\theta \left[\tilde{G}_T(\theta) | \mathcal{F}_T \right] = G_T(\theta).$$

Let $H(\theta)$ be an $\mathcal{F}_T$ measurable matrix and let δ_k be a sequence of positive real numbers converging to zero such that $\sum_{k=1}^{\infty} \delta_k = \infty$. Define the sequence $\{\hat{\theta}_k\}$ by the recursion

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \delta_k H(\hat{\theta}_k) \tilde{G}_{T,k}(\hat{\theta}_k), \tag{5}$$

where $\tilde{G}_{T,k}(\theta)$ are independent copies of $\tilde{G}_T(\theta)$. Sufficient conditions for the sequence (5) to converge to θ^*, the solution of $G_T(\theta) = 0$ are given in Spall (2003, pg 107). Of these conditions, the most important is that θ^* is an asymptotically stable equilibrium point of the ordinary differential equation

$$\frac{d\Upsilon_\tau}{d\tau} = -H(\Upsilon_\tau) G_T(\Upsilon_\tau). \tag{6}$$

The choice of $H(\theta)$ that makes θ^* an asymptotically stable equilibrium point of (6) can be difficult. For the optimal martingale estimating equation a good choice for $H(\theta)$ is the inverse of its quadratic characteristic, $\langle G(\theta) \rangle_T^{-1}$. In our simulation study we let $H(\theta)$ be an approximation of $\langle G(\theta) \rangle_T^{-1}$ obtained by approximating $\mathbb{E}_\theta \left(m_t(\theta) m_t(\theta)' | \mathcal{F}_{t-1} \right)$ using the OU process.

3 Estimating functions from the OU approximation

The optimal form of $a_t(\theta)$ given in equation (1) depends on quantities that are not available in closed form. Accurate numerical evaluation of $a_t^*(\theta)$ is computationally expensive as it involves numerical differentiation and evaluation of high order moments. Furthermore, numerical instability in evaluating the estimating function can lead to difficulties in solving the resulting estimating equation. It is therefore of interest to develop computationally inexpensive approximations to $a_t^*(\theta)$. We propose to use approximations of $a_t^*(\theta)$ based on a diffusion approximation to the queue length process.

The literature on queuing processes contains a number of diffusion approximations that are valid under a variety of conditions (Iglehart, 1965; Chen and Ye, 2001; Anisimov, 2002; Ward and Glynn, 2003). While these approximations could serve as a suitable starting point for constructing estimating equations, we take as our starting point the class of density dependent Markov processes and their approximation by the Ornstein-Uhlenbeck process.

Kurtz (1970) introduced the concept of density dependence for Markov processes in his examination of solutions to ordinary differential equations as approximations of Markov pure jump processes. Let f be a function mapping $\mathbb{R}^d \times \mathbb{Z}^d$ to $\mathbb{R}$ such that for each $l \in \mathbb{Z}^d,\ l \neq 0$, the function $f(x, l)$ is continuous in x. A one parameter family of Markov chains $X_c(t),\ c > 0$, is said to be density dependent if the transition rates have the form

$$q(m, m + l) = cf(c^{-1}m, l), \quad l \neq 0,\ m \in E.$$

In our study, the parameter c can be identified with the number of servers in the system. In a series of papers Kurtz (1970, 1971) and Barbour (1974, 1976, 1980) studied the asymptotic ($c \to \infty$) behaviour of this class of Markov processes (see also Either and Kurtz, 2005, chapter 11). Our construction of estimating functions uses only a few of their results which we now summarise.

Assume that the functions $f(x, l)$ are Lipshitz continuous in x for each l and that $f(x, l) \equiv 0$ for all but finitely many l. Define the function $F(x) := \sum_{l\neq 0} lf(x, l)$. These assumptions satisfy the conditions of Theorem 2.1 of Either and Kurtz (2005, chapter 11). Therefore, if $c^{-1}X_c(0) \to x^*$, almost surely, and if $F(x^*) = 0$, then for every $t \geq 0$,

$$\lim_{c\to\infty} \sup_{s\leq t} \left| c^{-1}X_c(s) - x^* \right| = 0,$$

almost surely.

Now define $Z_c(t) = c^{1/2}(c^{-1}X_c(t) - x^*)$. Assume that the Jacobian of F is continuous in a neighbourhood of x^*. Let A be the Jacobian of F evaluated at

x^* and let σ be the $d \times r$ matrix whose column vectors are the non-zero vectors $l\sqrt{f(x^*, l)}$. If $Z_c(0) \to z_0$, almost surely, then the conditions of Theorem 2.3 of Either and Kurtz (2005, chapter 11) are satisfied and $Z_c(t)$ converges weakly in $D[0, T]$ to the solution of the stochastic differential equation

$$dZ(t) = AZ(t)dt + \sigma dW(t), \quad Z(0) = z_0,$$

where $W(t)$ is r dimensional standard Brownian motion. The process $Z(t)$ is the OU process.

The conditional moments of the OU process provide a reasonable approximation to the conditional moments of $X(t)$. The conditional mean and conditional variance of the OU process can be evaluated using standard results (see Karatzas and Shreve, 1999, section 5.6). Let

$$\phi(t, z) = \mathbb{E}(Z(t)|Z(0) = z)$$

and let

$$\psi(t, x) = \mathbb{E}((Z(t) - \phi(t, x))(Z(t) - \phi(t, x))^T|Z(0) = z).$$

The functions $\phi(t, z)$ and $\psi(t, z)$ are given by the solutions to the initial value problems

$$\frac{d\phi(t, z)}{dt} = A\phi(t, z), \quad \phi(0, z) = z, \tag{7}$$

and

$$\frac{d\psi(t, z)}{dt} = A\psi(t, z) + \psi(t, z)A^T + \sigma\sigma^T, \quad \psi(0, z) = 0. \tag{8}$$

Note that $\psi(t, z)$ does not depend on z. Since the OU process is Gaussian, any higher order moments can be expressed as a function of the conditional mean and conditional variance. The following subsections provide details on the use of the OU approximation in formulating the estimating function for three types of queues. In each case, the sampling times are assumed to be the sequence of integers $1, 2, \ldots, T$. The results can easily be modified to allow for non-uniform sampling of the queue length process.

3.1 M/M/c queue

Consider an $M/M/c$ queue that has a Poisson arrival process with rate λ_c and c servers each having exponentially distributed service times with rate μ. The queue length process $X(t)$ is a Markov pure jump process on $\{0, 1, 2, \ldots\}$ with transition rates

$$
\begin{aligned}
q(m, m+1) &= \lambda_c, \\
q(m, m-1) &= \mu(m \wedge c).
\end{aligned}
$$

Assume that $\lim_{c\to\infty} c^{-1}\lambda_c = \lambda^*$. The OU approximation of $X(t)$ as $c \to \infty$ was first derived by Iglehart (1965). Applying the above results, we find $F(x) = \lambda^* - \mu(x \wedge 1)$. If we assume that $\lambda_c < c\mu$ then the parameters of the OU process approximating $X(t)$ are $x^* = \lambda^*/\mu$, $A = -\mu$ and $\sigma\sigma^T = 2\lambda^*$. The conditional mean and conditional variance of $X(t)$ are approximated by

$$
\begin{aligned}
\tilde{M}(x; \theta) &= \mu^{-1}\lambda_c + e^{-\mu}(x - \mu^{-1}\lambda_c), \\
\tilde{V}(\theta) &= \mu^{-1}\lambda_c(1 - e^{-2\mu}).
\end{aligned}
$$

To form the estimating function we first need to choose a martingale difference sequence. It is known that the score function of a Gaussian process is a quadratic function of the data. So that our class of estimating functions is close to the true score function, at least when c is large, we propose to use the martingale difference sequence based on $(X(t), X^2(t))$. That is,

$$
m_t(\theta) = \Big[X_t - M(X_{t-1}; \theta) \quad (X_t - M(X_{t-1}; \theta))^2 - V(X_{t-1}; \theta)\Big]'.
$$

Using this martingale difference sequence, the optimal estimating function is obtained by setting

$$
a_t^*(\theta) = [-\partial M(X_{t-1}; \theta)' \quad -\partial V(X_{t-1}; \theta)'] \begin{bmatrix} V(X_{t-1}; \theta) & S(X_{t-1}; \theta) \\ S(X_{t-1}; \theta) & K(X_{t-1}; \theta) \end{bmatrix},
$$

where ∂ denotes the partial derivative with respect to θ,

$$
S(x; \theta) = \mathbb{E}_\theta[(X_t - M(x; \theta))^3 | X_{t-1} = x]
$$

and

$$
K(x; \theta) = \mathbb{E}_\theta[(X_t - M(x; \theta))^4 | X_{t-1} = x] - V(x; \theta).
$$

Applying the OU approximation, $S(x; \theta) \approx 0$ and $K(x; \theta) \approx 2\tilde{V}^2(\theta)$. The proposed estimating function is

$$
\begin{aligned}
&\sum_{t=1}^{T} \partial\tilde{M}(X_{t-1}; \theta)'\tilde{V}(\theta)^{-1} (X_t - M(X_{t-1}; \theta)) \\
&+ \sum_{t=1}^{T} \partial\tilde{V}(\theta)' \left(2\tilde{V}^2(\theta)\right)^{-1} \left((X_t - M(X_{t-1}; \theta))^2 - V(X_{t-1}; \theta)\right).
\end{aligned}
$$

Note that the functions M, V appearing in the estimating function are determined by simulation.

3.2 A priority queue with preemptive service

We now consider a queue with two priority classes. Customers of the high priority class arrive in the queue according to a Poisson process with rate $\lambda_{1,c}$. Customers of the low priority class arrive in the queue according to a Poisson process with rate $\lambda_{2,c}$ independent of the high priority class customers. The customers are served by c servers having independent exponentially distributed service times with rate μ. Customers of the high priority class are served before those of the low priority class. Furthermore, the queue is preemptive so that a high priority customer will displace a low priority customer that is being served. The low priority customer is returned to the queue. Formally, the queue length process $X(t) = (X_1(t), X_2(t))$ counting the number of customers of each priority class in the queue is a Markov pure jump process with transition rates

$$\begin{aligned}
q(X, X + e_1) &= \lambda_{1,c}, \\
q(X, X + e_2) &= \lambda_{2,c}, \\
q(X, X - e_1) &= \mu(X_1 \wedge c), \\
q(X, X - e_2) &= \mu(X_2 \wedge (0 \vee c - X_1)),
\end{aligned}$$

where $e_1 = (1,0)'$ and $e_2 = (0,1)'$. Assume that $\lambda_{1,c} + \lambda_{2,c} < c\mu$ and that $\lim_{c\to\infty} \lambda_{i,c}/c = \lambda_i^*$. Applying the results on density dependent processes, we have

$$F(x) = \begin{pmatrix} \lambda_1^* - \mu(x_1 \wedge 1) \\ \lambda_2^* - \mu(x_2 \wedge (0 \vee 1 - x_1)) \end{pmatrix}$$

The parameters of the approximating OU process are

$$x^* = \begin{bmatrix} \lambda_1^*/\mu \\ \lambda_2^*/\mu \end{bmatrix}, \quad A = \begin{bmatrix} -\mu & 0 \\ 0 & -\mu \end{bmatrix}, \quad \sigma\sigma^T = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_2^* \end{bmatrix}.$$

Therefore, $X_1(t)$ and $X_2(t)$ are asymptotically independent processes.

As in the previous example, we construct the estimating function using a quadratic function of the data. The estimating function is formed from the martingale difference sequence

$$m_t(\theta) = \begin{bmatrix} X_{1,t} - M_1(X_{t-1};\theta) \\ X_{2,t} - M_2(X_{t-1};\theta) \\ (X_{1,t} - M_1(X_{t-1};\theta))^2 - V_{11}(X_{t-1};\theta) \\ (X_{1,t} - M_1(X_{t-1};\theta))(X_{2,t} - M_2(X_{t-1};\theta)) - V_{12}(X_{t-1};\theta) \\ (X_{2,t} - M_2(X_{t-1};\theta))^2 - V_{22}(X_{t-1};\theta) \end{bmatrix}.$$

The exact form for $a_t^*(\theta)$ is quite complicated. However, a great deal of simplification is achieved by using the OU process approximation. First, note that $V_{12}(x;\theta) \approx 0$, since $X_1(t)$ and $X_2(t)$ are asymptotically independent. It follows that

$$\begin{aligned} &\mathbb{E}_\theta\left[\partial m_t(\theta)|\mathcal{F}_{t-1}\right] \\ &\approx \left[-\partial M_1(X_{t-1};\theta) \quad -\partial M_2(X_{t-1};\theta) \quad -\partial V_{11}(X_{t-1};\theta) \quad 0 \quad -\partial V_{22}(X_{t-1};\theta)\right]. \end{aligned}$$

Secondly, the matrix $\mathbb{E}_\theta\left[m_t(\theta)m_t(\theta)'|\mathcal{F}_{t-1}\right]$ is approximately diagonal with non-zero entries

$$\left[V_{11}(X_{t-1};\theta) \;\; V_{22}(X_{t-1};\theta) \;\; 2V_{11}^2(X_{t-1};\theta) \;\; 2V_{11}(X_{t-1};\theta)V_{22}(X_{t-1};\theta) \;\; 2V_{22}^2(X_{t-1};\theta)\right]$$

Finally, we can approximate $M_i(x;\theta)$ and $V_{ii}(x;\theta)$ by

$$\begin{aligned} \tilde{M}_i(x;\theta) &= \mu^{-1}\lambda_{i,c} + e^{-\mu}(x_i - \mu^{-1}\lambda_{i,c}), \\ \tilde{V}_{ii}(\theta) &= \mu^{-1}\lambda_{i,c}(1 - e^{-2\mu}). \end{aligned}$$

We combine these approximations to yield the estimating function

$$\begin{aligned} &\sum_{t=1}^{T} \partial\tilde{M}(X_{t-1};\theta)'\tilde{V}(\theta)^{-1}\left(X_t - M(X_{t-1};\theta)\right) \\ &+ \sum_{t=1}^{T} \partial\tilde{V}(\theta)'\left(2\tilde{V}^2(\theta)\right)^{-1}\left((X_t - M(X_{t-1};\theta))^2 - V(X_{t-1};\theta)\right), \end{aligned}$$

where $\partial\tilde{M}(x;\theta) = [\partial M_1(X_{t-1};\theta) \quad \partial M_2(X_{t-1};\theta)]$, $\partial\tilde{V}(\theta) = [\partial\tilde{V}_{11}(\theta)\ \partial\tilde{V}_{22}(\theta)]$ and $\tilde{V}(\theta)$ is the diagonal matrix whose non-zero entries are $[\tilde{V}_{11}(\theta)\ \tilde{V}_{22}(\theta)]$. It is important to note that although $X_1(t)$ and $X_2(t)$ are asymptotically independent as $c \to \infty$, they are dependent for any finite c. The functions $M_i(X_t;\theta)$ and $V_{ii}(X_t;\theta)$ appearing in the estimating function are functions of the vector $(X_{1,t}, X_{2,t})$ and not just the respective $X_{i,t}$.

3.3 A tandem queue

As a final example, we consider a simple tandem queue. Customers arrive in the queue according to a Poisson process with rate λ_c. Customers are served at the first station by c servers having exponentially distributed service times with rate μ_1. After being served at the first station, customers proceed to a second station where they are served by c servers having exponentially distributed service times with rate μ_2. The queue length process comprising of the number of customers at each station is a Markov pure jump process with transition rates

$$\begin{aligned} q(X, X+(1,0)') &= \lambda_c, \\ q(X, X+(-1,1)') &= \mu_1(X_1 \wedge c), \\ q(X, X+(0,-1)') &= \mu_2(X_2 \wedge c) \end{aligned}$$

We assume that $\lim_{c\to\infty} \lambda_c/c = \lambda^*$ and that $\lambda < c \min_i \mu_i$. The queue length process is density dependent with

$$F(x) = \begin{pmatrix} \lambda^* - \mu_1(x_1 \wedge 1) \\ \mu_1(x_1 \wedge 1) - \mu_2(x_2 \wedge 1) \end{pmatrix}.$$

The parameters of the approximating OU process are

$$x^* = \begin{bmatrix} \lambda^*/\mu_1 \\ \lambda^*/\mu_2 \end{bmatrix}, \quad A = \begin{bmatrix} -\mu_1 & 0 \\ \mu_1 & -\mu_2 \end{bmatrix}, \quad \sigma\sigma^T = \begin{bmatrix} 2\lambda^* & -\lambda^* \\ -\lambda^* & 2\lambda^* \end{bmatrix}.$$

Solving the ordinary differential equations (7) and (8), we obtain the approximating conditional mean and conditional variance functions,

$$\begin{aligned} M_1(x;\theta) \approx \tilde{M}_1(x;\theta) &= \frac{\lambda_c}{\mu_1} + e^{-\mu_1}\left(x_1 - \frac{\lambda_c}{\mu_1}\right), \\ M_2(x;\theta) \approx \tilde{M}_2(x;\theta) &= \frac{\lambda_c}{\mu_2} + e^{-\mu_2}\left(x_2 - \frac{\lambda_c}{\mu_2}\right) + \frac{\mu_1}{\mu_2 - \mu_1}\left(x_1 - \frac{\lambda_c}{\mu_1}\right)\left(e^{-\mu_1} - e^{-\mu_2}\right), \\ V_{11}(x;\theta) \approx \tilde{V}_{11}(\theta) &= \frac{\lambda_c}{\mu_1}\left(1 - e^{-2\mu_1}\right), \\ V_{12}(x;\theta) \approx \tilde{V}_{12}(\theta) &= \frac{\lambda_c}{\mu_2 - \mu_1}\left(e^{-(\mu_1+\mu_2)} - e^{-2\mu_1}\right), \\ V_{22}(x;\theta) \approx \tilde{V}_{22}(\theta) &= \frac{-\mu_1\lambda_c}{(\mu_2 - \mu_1)^2}\left(e^{-\mu_1} - e^{-\mu_2}\right)^2 + \frac{2\lambda_c}{\mu_2}\left(1 - e^{-\mu_2}\right). \end{aligned}$$

For simplicity we only consider the estimating function based on the martingale difference sequence $X_t - M(X_{t-1};\theta)$. The estimating function is given by

$$\sum_{t=1}^{T} \partial \tilde{M}(X_{t-1};\theta)' \tilde{V}^{-1}(\theta) \left[X_t - M(X_{t-1};\theta) \right].$$

We expect that the resulting estimator will be less efficient than an estimator obtained from a quadratic estimating equation. However, it will still be consistent, assuming the regularity conditions are satisfied.

4 Simulations

4.1 M/M/c

As stated in the introduction, the problems with the OU likelihood estimator has motivated our proposal to use estimators based on MEEs. Our first example provides a comparison of the two approaches using simulated data. We take the same $M/M/c$ queues used in Ross et al. (2007) as the basis of the comparison. The three $M/M/c$ queues have the following parameters: (A) a shopping queue comprised of 5 servers with a service rate of 0.175 and arrival rate 0.75, (B) a small telecommunications queue comprised of 50 servers with a service rate of 1/6 and arrival rate 85/12 and (C) a large telecommunications queue comprised of 300 servers with a service rate 0.09 and arrival rate 25.

Ross et al. (2007) have noted that larger sample sizes do not improve the OU likelihood estimate. Although we do not expect this to be true for the MEE estimate, we use two different lengths of time series; one as in Ross et al. (2007) and another approximately ten times longer.

For each queue and length of time series, 100 time series were simulated and the two estimators applied. We solved the quadratic MEE of section 3.1 using the stochastic root finding algorithm described in section 2.3. The OU likelihood estimates were obtained using a quasi-Newton solver.

Tables 1 – 3 provide summaries of the estimator performance for each of the three queues and two lengths of time series. The MEE approach provides a considerable reduction in the mean squared error over the OU likelihood approach. This is almost entirely due to a reduction of bias. As expected the improvement is greater for larger sample sizes and is less when the OU process is a good approximation for the queue length process.

It has already been noted that the confidence sets for the approximate likelihood estimator have poor coverage. Using the same simulated data as for Tables 1 – 3, we investigate the observed coverage levels for 95% confidence sets for the MEE estimator and compared this with the OU estimator. The

Table 1
(A) Small shopping queue: Empirical bias and standard deviation of estimates from simulated queue. RMSE is the ratio of the mean squared error of the MEE estimate to the OU estimate.

n	Parameter	Method	Bias	Std Dev	RMSE
240	λ	OU	0.0268	0.0858	
		MEE	-0.0082	0.0959	1.1470
	μ	OU	0.0856	0.0322	
		MEE	-0.0061	0.0210	0.0569
2000	λ	OU	0.0523	0.0326	
		MEE	-0.0057	0.0352	0.3322
	μ	OU	0.1050	0.0188	
		MEE	-0.0029	0.0086	0.0072

Table 2
(B) Small telecommunication system: Empirical bias and standard deviation of estimates from simulated queue. RMSE is the ratio of the mean squared error of the MEE estimate to the OU estimate.

n	Parameter	Method	Bias	Std Dev	RMSE
240	λ	OU	0.1768	0.6234	
		MEE	0.0479	0.6558	1.0290
	μ	OU	0.0089	0.0176	
		MEE	0.0009	0.0176	0.5609
2000	λ	OU	0.1145	0.2727	
		MEE	-0.0338	0.2736	0.8674
	μ	OU	0.0078	0.0073	
		MEE	-0.0010	0.0062	0.3487

confidence sets for the MEE estimator are constructed using expression (4). The confidence sets for the OU estimator are similarly constructed using

$$\left\{\theta : u_T'(\theta) I_T(\theta)^{-1} u_T(\theta) \leq \chi^2_{p;(1-\alpha)}\right\}, \tag{9}$$

where $u_T(\theta)$ is the score (first derivative of the log likelihood), $I_T(\theta)$ is the observed information matrix (negative of the Hessian of the log likelihood) and $\chi^2_{p;(1-\alpha)}$ is the $1-\alpha$ percentile of the χ^2 distribution with p degrees of freedom. The results are reported in Table 4. Although 100 simulations is too

Table 3
(C) Large telecommunication system: Empirical bias and standard deviation of estimates from simulated queue. RMSE is the ratio of the mean squared error of the MEE estimate to the OU estimate.

n	Parameter	Method	Bias	Std Dev	RMSE
600	λ	OU	0.1103	1.7816	
		MEE	-0.0020	1.8044	1.0218
	μ	OU	0.0008	0.0065	
		MEE	-0.0002	0.0064	0.9798
6000	λ	OU	0.0818	0.5279	
		MEE	-0.0783	0.5218	0.9755
	μ	OU	0.0009	0.0020	
		MEE	-0.0003	0.0019	0.7756

Table 4
Observed coverage level for 95% confidence sets formed using expressions (4) and (9).

MEE				OU			
length	A	B	C	length	A	B	C
short	0.94	0.95	0.96	short	0.09	0.65	0.84
long	0.96	0.98	0.97	long	0.00	0.41	0.57

few to be able to detect small deviations from the theoretical coverage level, it is sufficient to detect the large deviations observed for the OU likelihood method. For the OU likelihood method, coverage of the confidence intervals improve as c increases and deteriorates as the sample size increases. This is expected as the bias in the OU likelihood estimates decrease as c increases. The confidence sets from the MEE method are possibly conservative due to the small amount of extra variation introduced by the approximation of the conditional mean and the conditional variance in the estimating equation using simulation.

4.2 A priority queue with preemptive service

We now examine the performance of the estimating equation approach as applied a queue with two priority classes. Two queues of the form described in section 3.2 were simulated. The first queue had arrival rates 0.3 and 0.45

Table 5
Empirical bias and standard deviation of estimates from simulated queues with two priority classes.

Queue One			Queue Two		
Parameter	Bias	Std Dev	Parameter	Bias	Std Dev
λ_1	-0.0077	0.0483	λ_1	0.0384	0.2743
λ_2	0.0068	0.0525	λ_2	0.0585	0.3083
μ	0.0055	0.0172	μ	0.0022	0.0125

for the high and low priority classes, respectively. There were 5 servers in the queue with a service rate of 0.175. The second queue had arrival rates of 3 and 4.083 for the high and low priority classes, respectively. There were 50 servers in the queue with a service rate of 1/6. The queues were simulated and 240 observations of queue length were recorded at unit time intervals. The estimate of θ from solving the estimating equation was also recorded. This was repeated 100 times for both queues to obtain empirical estimates of the bias and standard deviation of the estimator. The results are summarised in Table 5. We see that any bias in the estimators is smaller than could be detected by this simulation study. A good level of accuracy was achieved in both cases.

4.3 A tandem queue

Finally, we examine the performance of the estimating equation approach as applied to tandem queues with two stations. Two queues of the form described in section 3.3 were simulated. The first queue had an arrival rate of 0.75 and each station comprised 5 servers with a service rate 0.175. The second queue had arrival rate of 3.54 and each station comprised of 25 servers with a service rate of 1/6. The same procedure that was used for the priority queues was applied for the two tandem queues. The results are summarised in Table 6. Although there is a noticeable drop in efficiency due to using a linear instead of a quadratic estimating equation, the estimator still appears to perform well. There also appears to be a positive bias in the estimates of the arrival and service rates.

Table 6
Empirical bias and standard deviation of estimates from simulated tandem queues with two stations.

Queue One			Queue Two		
Parameter	Bias	Std Dev	Parameter	Bias	Std Dev
λ	0.1197	0.1572	λ	0.1212	0.3882
μ_1	0.0275	0.0343	μ_1	0.0059	0.0177
μ_2	0.0260	0.0327	μ_2	0.0057	0.0182

5 Discussion

The MEE approach yields improved estimators compared to the OU likelihood approach. Unlike the OU likelihood approach, the performance of the MEE estimators does not critically depend on the closeness of the queue length process to an OU process. The MEE estimators typically have smaller mean squared error due to a large reduction in the bias. Furthermore, the confidence sets constructed from the MEE appear to have better coverage. Although there is considerable support for preferring the MEE approach, we need to acknowledge two important facts. Firstly, we have assumed that the queuing process satisfies certain regularity conditions so that the MEE estimator would be consistent and asymptotically normal. These conditions still need to be verified. Secondly, these improvements are obtained at the cost of additional computations.

We have used the approximating OU process as a means of approximating the optimal MEE. It may not always be possible to use the OU approximation in this way, especially if the OU approximation does not depend on one or more of the process parameters. An alternative approach would be to approximate the required conditional moments using a small sampling time approximation. In this case, the conditional distribution of $X(t+h)$ given $X(t) = x$, and hence the conditional moments, is approximated assuming h is small. Truncation of the uniformization expansion is one approach to obtain this approximation.

Finally, we note that it is possible to modify the ideas of this paper to deal with non-Markovian queues. Suppose that we have an approximate likelihood function for the queue length process. Let $l(\theta)$ denote the approximate log likelihood function. An unbiased estimating equation is given by

$$\frac{\partial l}{\partial \theta}(\theta) - \mathbb{E}_\theta \left(\frac{\partial l}{\partial \theta}(\theta) \right) = 0.$$

If $l(\theta)$ is exactly the log likelihood of the queue length process then the above expectation is known to be zero and we obtain the maximum likelihood estimate. In general, the above expectation does not have a simple form but can be evaluated numerically if the queue can be simulated.

References

Anisimov, V.V., Diffusion approximation in overloaded switching queuing models, Queueing Syst., 40, 143-182 (2002)

Ausín, M.C., Wiper, M.P., and Lillo, R.E., Bayesian estimation for the $M/G/1$ queue using a phase-type approximation, J. Statist. Plann. Inference, 118, 83-101 (2004)

Barbour, A.D., On a functional central limit theorem for Markov population processes, Adv. Appl. Probab., 6, 21-39 (1974)

Barbour, A.D., Quasi-stationary distributions in Markov population processes, Adv. Appl. Probab., 8, 296-314 (1976)

Barbour, A.D., Equilibrium distributions in Markov population processes, Adv. Appl. Probab., 591-614 (1980)

Basawa, I.V., Bhat, U.N., and Zhou, J., Parameter estimation using partial information with applications to queuing and related models, Statist. Probab. Lett., 78, 1375-1383 (2008)

Bibby, B.M., Sørensen, M., Martingale estimation functions for discretely observed diffusion processes, Bernoulli, 1, 17-39 (1995)

Bibby, B.M., Jacobsen, M. and Sørensen, M., Estimating functions for discretely sampled diffusion-type models, MaPhySto Research Report, no. 19 (2004) downloaded from `www.maphysto.dk/publications2/MPS-RR/2004/19.pdf` 20/07/2011.

Blum, M.G.B., and François, O., Non-linear regression models for approximate Bayesian computation, Stat. Comput., 20, 63-73 (2010)

Chen, H., and Ye, H.Q., Existence condition for the diffusion approximations of multiclass priority queuing networks, Queueing Syst., 38, 435-470 (2001)

Either, S.N., and Kurtz, T.G., Markov processes: characterization and convergence, Wiley Interscience, Hoboken (2005)

Fearnhead, P., Filtering recursions for calculating likelihoods for queues based on inter-departure time data, Stat. Comput., 14, 261-266, (2004)

Heggland, K., and Frigessi, A., Estimating functions in indirect inference, J. R. Stat. Soc. Ser. B, 66, 447-462 (2004)

Heyde, C.C., Quasi-likelihood and its application, Springer, New York (1997).

Heyde, C.C. and Morton, R. Quasi-likelihood and generalizing the EM algorithm, J. R. Statist. Soc. B, 58, 317-327 (1996)

Iglehart, D.L., Limiting diffusion approximations for the many server queue and the repairman problem, J. Appl. Probab., 2, 429-441 (1965)

Karatzas, I., and Shreve, S.E., Brownian motion and stochastic calculus, 2nd edition, Springer, New York (1999).

Kurtz, T.G., Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Probab., 7, 49-58 (1970)

Kurtz, T.G., Limit theorems for sequences of jump Markov processes approximating ordinary differential processes, J. Appl. Probab., 8, 344-356 (1971)

Novak, A., and Watson, R., Determining an adequate probe separation for estimating the arrival rate in an $M/D/1$ queue using single-packet probing, Queueing Syst., 61, 255-272 (2009)

Ross, J.V., Taimre, T., and Pollett, P.K., Estimation for queues from queue length data, Queueing Syst., 55, 131-138 (2007)

Spall, J.C., Introduction to stochastic search and optimization, Wiley, New Jersey (2003)

Tan, X., and Knessl, C., A fork-join queuing model: Diffusion approximation, integral representations and asymptotics, Queueing Syst., 22, 287-322 (1996)

Ward, A.R., and Glynn, P.W., A diffusion approximation for a Markovian queue with reneging, Queueing Syst., 43, 103-128 (2008)