On the modified-offered-load approximation for the nonstationary Erlang loss model

W. A. Massey and W. Whitt

Mathematical Sciences Research Center, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636, U.S.A.

Based on the behavior of real telephone systems, a natural model to use is the nonstationary Erlang loss model, i.e. the $M_t/G/s/0$ queue, which has $s$ servers, no extra waiting space and a nonhomogeneous Poisson arrival process ($M_t$). In 1975 Jagerman proposed the modified-offered-load (MOL) approximation for this model. With MOL, the distribution of the number of busy servers in the $M_t/G/s/0$ queue at time $t$ is approximated by the steady-state distribution of the stationary $M/G/s/0$ queue with an offered load (arrival rate times mean service time) equal to the mean number of busy servers in the $M_t/G/\infty$ queue at time $t$. In this paper, we find an expression for the difference between the $M_t/G/s/0$ queue length distribution and its MOL approximation. From this expression we extract bounds on the error and deduce when one distribution stochastically dominates the other.

1 Introduction

The probabilistic modelling of the number of busy lines in telephone trunk groups is one of the fundamental problems that led to the development of queueing theory. It was first formulated as an $M/G/s/0$ queue by Erlang [6]. He gave an exact solution for the steady state distribution, which gave rise to the well known Erlang blocking formula. This formula states that if $Q(t)$ is the random queue length at time $t$ (queueing here means “waiting” for service completion), then

$$\lim_{t \to \infty} P(Q(t) = s) = \beta(\lambda/\mu, s) \equiv \frac{(\lambda/\mu)^s}{s!} \sum_{k=0}^{s} \frac{(\lambda/\mu)^k}{k!},$$

(1.1)

where $\lambda$ is the Poisson arrival rate, $1/\mu \equiv E[S]$ is the mean of the random service time $S$ and $s$ equals the total number of servers (trunk lines). Since Poisson arrivals see time averages, $\beta(\lambda/\mu, s)$ is also the long-run proportion of arrivals that are lost.

For applications, a restrictive assumption in the $M/G/s/0$ model is the constant arrival rate. Significant steps were made to solve this problem starting in the 1930's, see Palm [15], Khintchine [10], and Prékopa [16]. They found the exact solution for the time-dependent distribution in the $M_t/G/\infty$ model. This infinite-server queue captures the effect of a time varying mean arrival rate and general service times, but at the expense of letting the total
number of servers be infinite. If $Q^\star(t)$ equals the queue length at time $t$ in the $M_t/G/\infty$ model, and $Q^\star(t_0) = 0$ for some $t_0 < t$, then for all non-negative integers $k$,

\[
P(Q^\star(t) = k) = e^{-m^\star(t)} \frac{m^\star(t)^k}{k!} \quad \text{where} \quad m^\star(t) = E\left[\int_{t-s}^{t} \lambda(\tau)d\tau\right], \tag{1.2}
\]

with $\lambda(t) = 0$ for all $t < t_0$. A simple direct approximation for the blocking probability $P(Q(t) = s)$ in the $M_t/G/s/0$ model is the tail probability $P(Q^\star(t) \geq s)$.

These exact solutions to the $M/G/s/0$ and $M_t/G/\infty$ models led to a better technique for approximating the time-dependent queue length distribution in the $M_t/G/s/0$ model, called the the modified-offered-load approximation (MOL), see Jagerman [7]. Since the Erlang blocking formula is a function of $\lambda/\mu$, and $\lambda/\mu$ is the mean steady state queue length in the stationary $M/G/\infty$ queue, we should obtain a reasonable approximation for the time-dependent blocking probability in the $M_t/G/s/0$ queue if we substitute $m^\star(t)$ for $\lambda/\mu$ in the Erlang blocking formula. Thus the MOL approximation is

\[
P(Q(t) = s) \approx \beta(m^\star(t), s) = P(Q^\star(t) = s | Q^\star(t) \leq s), \tag{1.3}
\]

where $\beta$ is given by (1.1) and $m^\star(t)$ is given by (1.2).

We applied the MOL approximation to help understand the impact of the service-time distribution in an $M_t/G/s/0$ queue in Davis et al. [2]. The MOL approximation was a major motivation for our papers with Eick [4], [5] on the $M_t/G/\infty$ model. Moreover, since a solution exists for the transient distribution of the $M_t/G_t/\infty$ queue, see Brown and Ross [1] and Massey and Whitt [14], we can apply the MOL approximation to the $M_t/G_t/s/0$ queue as well.

Our goal here is to create a mathematical theory supporting this heuristic approximation. In Section 4, we do so by constructing a formal solution to the error between the exact probability solution and the MOL approximation for the case of time-dependent phase-type service. From this main result, we derive simple, computable error bounds for MOL. For the $M_t/M/s/0$ queue, we obtain the bound

\[
\sup_{0 \leq s \leq t} |P(Q(\tau) = s) - \beta(m^\star(\tau), s)| \leq 2 \int_{0}^{t} \beta(m^\star(\tau), s)(1 - \beta(m^\star(\tau), s)) \left| \frac{dm^\star}{d\tau}(\tau) \right| d\tau, \tag{1.4}
\]

where we assume that the distribution of $Q(0)$ is the steady-state $M/M/s/0$ distribution with parameter $m^\star(0)$, which is a family of distributions that includes the point masses at 0 and $s$. For the more general $M_t/G/s/0$ system, we obtain the bound

\[
\sup_{0 \leq s \leq t} |P(Q(\tau) = s) - \beta(m^\star(\tau), s)| \leq 2 \int_{0}^{t} \beta(m^\star(\tau), s) \left| \frac{dm^\star}{d\tau}(\tau) \right| d\tau. \tag{1.5}
\]

These error bounds imply that the MOL approximation is asymptotically correct as either the derivative of $m^\star(t)$ or the tail probability $P(Q^\star(t) \geq s)$ in the $M_t/G_t/\infty$ model approaches 0. In turn, these limits for the $M_t/G/\infty$ model hold as the derivative of $\lambda(t)$ approaches 0 and as $s \to \infty$. More generally, these bounds support the intuition that MOL should perform better when the arrival rate $\lambda(t)$ changes more slowly and when the blocking probability is lower.

For brevity, we omit all proofs here; they appear in an expanded version of this paper.
2 The $M_t/PH_t/s/0$ Queue

We define the $M_t/PH_t/s/0$ queueing system as follows. It has $s$ separate servers, each with a common time-dependent phase-type service, and an arrival process that is non-homogeneous Poisson. The class of phase-type service-time distributions is quite general, because phase-type distributions are dense in the space of all distributions. Restricting attention to phase-type service time distributions enables us to construct an extended finite state space such that the queue length process is Markovian in continuous time. Let $C$ equal the finite set of service phases (which we assume does not change with time). To obtain a general state description that makes our system Markovian, we count the number of customers in each phase of service. We define $S_C$ to be the corresponding state space, allowing arbitrary numbers of customers. The states in $S_C$ can be denoted by $k$, where every $k \in S_C$ is written as the formal sum $k = \sum_{\alpha \in C} k_\alpha e_{\alpha}$, such that $e_{\alpha}$ is an independent basis vector, corresponding to the service phase $\alpha$, and each $k_\alpha$ is a non-negative integer, representing the number of customers in service phase $\alpha$. The set $S_C$ is the state space for the case of $s = \infty$. In algebraic terms, $S_C$ is referred to as the free abelian semigroup generated by the set $C$, in contrast to the free nonabelian semigroup structure used in Massey [12] for the state space of a multiclass single server queue. Finally, if we denote the length of $k$ as $|k|$, which equals $\sum_{\alpha \in C} k_\alpha$, then the state space for our queueing model $M_t/PH_t/s/0$, will be $S_C(s)$, where $S_C(s) = \{ k | k \in S_C \text{ and } |k| \leq s \}$.

Now let $\{ Q(t) | t \geq 0 \}$ be the Markovian queue length process with state space $S_C(s)$. Its infinitesimal generator will be constructed from the following parameters:

- $\lambda_\alpha(t) = \text{the external arrival rate at time } t \text{ for a customer that initiates service in phase } \alpha.$
- $\mu_\alpha(t) = \text{the service rate at time } t \text{ for phase } \alpha.$
- $\rho_{\alpha\beta}(t) = \text{the probability that phase } \beta \text{ service is initiated at time } t, \text{ given that phase } \alpha \text{ service has just terminated.}$
- $q_\alpha(t) = \text{the probability that the entire service has terminated at time } t, \text{ given that phase } \alpha \text{ service has just terminated.}$

If $p(k, t) \equiv P(Q(t) = k)$, then for $|k| < s$, $Q(t)$ has the following set of forward equations:

$$
\frac{dp(k, t)}{dt} = \sum_{\alpha \in C} \left[ \lambda_\alpha(t) \text{sgn}(k_\alpha) p(k - e_\alpha, t) + \mu_\alpha(t)(k_\alpha + 1) q_\alpha(t) p(k + e_\alpha, t) + \sum_{\beta \in C} \mu_\beta(t) (k_\beta + 1) \rho_{\alpha\beta}(t) \text{sgn}(k_\alpha) p(k - e_\alpha + e_\beta, t) - (\lambda_\alpha(t) + \mu_\alpha(t) k_\alpha) p(k, t) \right],
$$

where $\text{sgn}(k)$ equals 0 if $k = 0$, and 1 if $k > 0$. When $|k| = s$, we have

$$
\frac{dp(k, t)}{dt} = \sum_{\alpha \in C} \left[ \lambda_\alpha(t) \text{sgn}(k_\alpha) p(k - e_\alpha, t) + \sum_{\beta \in C} \mu_\beta(t) (k_\beta + 1) \rho_{\alpha\beta}(t) \text{sgn}(k_\alpha) p(k - e_\alpha + e_\beta, t) - \mu_\alpha(t) k_\alpha p(k, t) \right].
$$
Letting \( \ell(S_C(s)) \) be the vector space for real valued functions on \( S_C(s) \), we can encode these equations as

\[
\frac{d}{dt} p(t) = p(t) A(t) \quad \text{where} \quad p(t) = \sum_{k \in S_C(s)} P(Q(t) = k)e_k,
\]

and \( A(t) \) is the corresponding infinitesimal generator that is a linear operator on \( \ell(S_C(s)) \) composed of the arrival and service rates for the queueing process. The \( e_k \)'s are the unit basis vectors for \( \ell(S_C(s)) \), where each \( e_k \) corresponds to the indicator function for the singleton set \{k\}. In general, \( p(t) \) is a probability vector, since it is a vector encoding of the probability distribution given by \( p(k, t) \). We will use the terms probability vector and probability distribution interchangeably. Formally, we can solve for \( p(t) \), and get

\[
p(t) = p(0) E_A(t),
\]

where \( E_A(t) \) is the time-ordered exponential of the family of generators \( \{ A(\tau) \mid 0 \leq \tau \leq t \} \).

When \( A \) is a constant operator, then the corresponding time-ordered exponential is just \( \exp(t A) \). In general, it is the unique operator solution to the equation

\[
\frac{d}{dt} E_A(t) = E_A(t) A(t),
\]

where \( E_A(0) = I \), the identity operator. For all \( 0 \leq \tau \leq t \), it will also be useful to define

\[
E_A(\tau, t) = E_A(\tau)^{-1} E_A(t).
\]

A thorough treatment of the issues of existence, uniqueness, and construction of time-ordered exponentials can be found in Dollard and Friedman [3].

### 3 The \( M_t/PH_t/\infty \) Queue

Our approximate analysis of the \( M_t/PH_t/s/0 \) employs the exact solution for its infinite-server counterpart, the \( M_t/PH_t/\infty \) queue. Let \( \{ Q^*(t) \mid t \geq 0 \} \) be the \( M_t/PH_t/\infty \) queue length process. Its marginal probabilities \( q(k, t) = P(Q^*(t) = k) \) for all \( k \in S_C \), will then solve the following set of forward equations:

\[
\frac{d}{dt} q(k, t) = \sum_{\alpha \in C} \left[ \lambda_\alpha(t) \text{sgn}(k_\alpha) q(k - e_\alpha, t) + \mu_\alpha(t) (k_\alpha + 1) q_\alpha(t) q(k + e_\alpha, t) \right]
+ \sum_{\beta \in C} \mu_\beta(t) (k_\beta + 1) p_\beta(t) \text{sgn}(k_\beta) q(k - e_\beta + e_\alpha, t) - (\lambda_\alpha(t) + \mu_\alpha(t) k_\alpha) q(k, t) .
\]

Now for any \( x \) in \( \ell(C) \), the vector space of real-valued functions on \( C \), and any state \( k \in S_C \), define the following useful operations:

\[
x^k = \prod_{\alpha \in C} x_\alpha^{k_\alpha}, \quad k! = \prod_{\alpha \in C} k_\alpha!, \quad \text{and} \quad |x| = \sum_{\alpha \in C} |x_\alpha| .
\]
where $x_\alpha = x(\alpha)$. We will also represent $x$ by the formal sum $\sum_{\alpha \in C} x_\alpha e_\alpha$. Hence $|x|$ is the $\ell_1$-norm applied to $x$. In this notation, the multinomial theorem is transformed into

$$
\sum_{|k|=s} \frac{x^k}{k!} = \frac{|x|^s}{s!}.
$$

(3.2)

Theorem 8.2 of Massey and Whitt [14] gives the exact solution for the $M_t/PH_t/\infty$ queue with appropriate initial distributions, as

$$
q(k,t) = \frac{e^{-m^*(t)}m^*(t)^k}{k!},
$$

(3.3)

where $m^*(t) = \sum_{\alpha \in C} m^*_\alpha(t)e_\alpha$ and $m^*(t) = |m^*(t)| = \sum_{\alpha \in C} m^*_\alpha(t)$, such that the $m^*_\alpha(t)$'s solve the set of differential equations

$$
\frac{d}{dt} m^*_\alpha(t) = \lambda_\alpha(t) + \sum_{\beta \in C} \mu_\beta(t)m^*_\beta(t)p_{\beta\alpha}(t) - \mu_\alpha(t)m^*_\alpha(t)
$$

(3.4)

for all $\alpha \in C$, with arbitrary $m^*(0)$. The solution (3.3) is valid provided that the initial distribution $p(k,0)$ is also of the same form depending on the initial mean vector $m^*(0)$.

### 4 The Fundamental Identity and Bounds for MOL

The MOL approximation is defined to be $p^*(k,t)$ for $S_C$, where

$$
P(Q(t) = k) \approx p^*(k,t) \equiv \frac{m^*(t)^k}{k!} \left/ \sum_{j=0}^{s} \frac{m^*(t)^j}{j!} \right. = P(Q^*(t) = k \mid |Q^*(t)| \leq s),
$$

(4.1)

where the components of the vector $m^*(t) = \sum_{\alpha \in C} m^*_\alpha(t)e_\alpha$ solve the differential equations given by (3.4), with arbitrary initial vector $m^*(0)$. We now present our fundamental result:

**Theorem 4.1** Let $\{Q(t) \mid t \geq 0\}$ be the Markovian queueing process for $M_t/PH_t/s/0$ with the family of infinitesimal generators $\{A(t) \mid t \geq 0\}$. Let $p(t)$ be the probability vector for the distribution of $Q(t)$, with an initial distribution $p(0) = p^*(0)$, which is of the form (3.3) for arbitrary $m^*(0)$. Let $p^*(t)$ be the probability vector for the modified-offered-load approximation, then

$$
p^*(t) - p(t) = \sum_{|k|=s} \int_0^t p^*(k,\tau) \cdot (p^*(\tau) - e_k) E_A(\tau, t) \, dm^*(\tau),
$$

(4.2)

where $E_A(\tau, t)$ is given by (2.4), the signed measure $dm^*(\tau)$ is formally the derivative of $m^*$ times $d\tau$, and

$$
dm^*(\tau) = \left( \sum_{\alpha \in C} \lambda_\alpha(\tau) - \mu_\alpha(\tau)m^*_\alpha(\tau)q_\alpha(\tau) \right) d\tau.
$$

(4.3)

We now apply Theorem 4.1 to obtain bounds and inequalities. First, we obtain bounds by simply bounding the time-ordered exponential $E_A(\tau, t)$ in (4.2) by 1.
Corollary 4.2 In the setting of Theorem 4.1, We have the following bounds for the error due to the modified-offered-load approximation:

\[
\sup_{0 \leq \tau \leq t} |p^*(\tau) - p(\tau)| \leq 2 \cdot \sum_{|k|=s} \int_0^t p^*(k, \tau)(1 - p^*(k, \tau)) \, |dm^*| (\tau) \tag{4.4}
\]

\[
\leq 2 \cdot \int_0^t \beta(m^*(\tau), s) \left(1 - \frac{\beta(m^*(\tau), s)}{(l+1)^{c-1}} \right) \, |dm^*| (\tau) \tag{4.5}
\]

\[
\leq 2 \cdot \int_0^t \beta(m^*(\tau), s) \, |dm^*| (\tau), \tag{4.6}
\]

where $|\cdot|$ as applied to $p^*(\tau)$ and $p(\tau)$ is the $\ell_1$-norm on $\ell(S_C(s))$, the measure $|dm^*| (\tau)$ is

\[
|dm^*| (\tau) \equiv \left| \frac{dm^*}{d\tau} (\tau) \right| \, d\tau, \quad \text{and} \quad \beta(m^*(t), s) = \sum_{|k|=s} p^*(k, t).
\]

If $Q(0)$ has a distribution that is not of the form (4.1), then we can construct a process $\tilde{Q}$ that has the same infinitesimal generator, but an initial distribution of the proper form. We then have

\[
\sup_{0 \leq \tau \leq t} |p^*(\tau) - p(\tau)| \leq |p(0) - \tilde{p}(0)| + \sup_{0 \leq \tau \leq t} |p^*(\tau) - \tilde{p}(\tau)|, \tag{4.8}
\]

where $\tilde{p}$ is the probability vector for $\tilde{Q}$, and now the above corollary applies.

5 MOL Bounds for the $M_t/M_t/s/0$ Queue

Now we restrict ourselves to one class or $|C| = 1$, which gives us the $M_t/M_t/s/0$ queue. It follows that $S_C(s) = \{0, 1, \ldots, s\}$, which is a totally ordered set. Moreover, our fundamental relation (4.2) simplifies to

\[
p^*(t) - p(t) = \int_0^t \beta_s(m^*(\tau)) \cdot (p^*(\tau) - e_s) \, E_A(\tau, t) \, dm^* (\tau). \tag{5.1}
\]

The next proposition establishes a stochastic comparison between the $M_t/M_t/s/0$ queue and its MOL approximation. We say that a probability vector $p_1$ is stochastically dominated by $p_2$, and write $p_1 \leq_{st} p_2$, if

\[
\sum_{j=k}^s p_1(j) \leq \sum_{j=k}^s p_2(j) \quad \text{for all } k = 0, 1, \ldots, s. \tag{5.2}
\]

In terms of operators and componentwise ordering of vectors, $p_1 \leq_{st} p_2$ is equivalent to $p_1 K \leq p_2 K$, where $K = (I - L)^{-1}$ with $L$ equalling the left shift operator on row vectors, or

\[
K = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
1 & 1 & 0 & \cdots & 0 & 0 \\
1 & 1 & 1 & \cdots & 0 & 0 \\
& \cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 1 & 1 & \cdots & 1 & 0 \\
1 & 1 & 1 & \cdots & 1 & 1 \\
\end{bmatrix}. \tag{5.3}
\]
Proposition 5.1 For the $M_t/M_t/s/0$ system, if $m^*(0) \leq \lambda(0)/\mu(0)$ and $\lambda/\mu$ is an increasing function on $[0, t]$, then the modified offered load distribution is stochastically dominated by the exact distribution for $Q$ on $[0, t]$, or

$$p^*(\tau) \leq \mu_p(\tau), \quad \text{for all } \tau \in [0, t].$$

(5.4)

In particular, $\beta_s(m^*)$ underestimates the actual blocking probability on $[0, t]$. Conversely, if $m^*(0) \geq \lambda(0)/\mu(0)$ and $\lambda/\mu$ is an decreasing function on $[0, t]$, then the exact distribution for $Q(t)$ is stochastically dominated by the the modified offered load distribution at time $t$ and $\beta_s(m^*)$ overestimates the actual blocking probability on $[0, t]$.

Theorem 5.2 For all $t \geq 0$ with $p(0) = p^*(0)$, we have

$$|(p^*(t) - p(t))\mathbf{K}| \leq \int_0^t \beta_s(m^*(\tau))(s - m^*(\tau)(1 - \beta_s(m^*(\tau)))) \exp \left(-\int_\tau^t \mu(r) dr\right) |dm^*(\tau)|.$$

(5.5)

Corollary 5.3 Using the same hypothesis as above, we have

$$|P(Q(t) = s) - \beta_s(m^*(t))|$$

$$\leq \int_0^t \beta_s(m^*(\tau))(s - m^*(\tau)(1 - \beta_s(m^*(\tau)))) \exp \left(-\int_\tau^t \mu(r) dr\right) |dm^*(\tau)|$$

(5.6)

and

$$|E[Q(t)] - m^*(t)(1 - \beta_s(m^*(t)))]$$

$$\leq \int_0^t \beta_s(m^*(\tau))(s - m^*(\tau)(1 - \beta_s(m^*(\tau)))) \exp \left(-\int_\tau^t \mu(r) dr\right) |dm^*(\tau)|.$$

(5.7)

Corollary 5.4 If $Q(0) = m^*(0) = 0$, $\lambda$ is a bounded function, and $\mu$ is a constant function, then

$$\sup_{t \geq 0} |(p^*(t) - p(t))\mathbf{K}| \leq \frac{s|\lambda'|_\infty}{\mu^2} \beta_s\left(\frac{|\lambda|_\infty}{\mu}\right),$$

(5.8)

where $|\lambda|_\infty$ and $|\lambda'|_\infty$ are respectively the supremum norm for $\lambda$ and its derivative for all $t \geq 0$. From this bound it follows that

$$\sup_{t \geq 0} |P(Q(t) = s) - \beta_s(m^*(t))| \leq \frac{s|\lambda'|_\infty}{\mu^2} \beta_s\left(\frac{|\lambda|_\infty}{\mu}\right)$$

(5.9)

and

$$\sup_{t \geq 0} |E[Q(t)] - m^*(t)(1 - \beta_s(m^*(t)))] \leq \frac{s|\lambda'|_\infty}{\mu^2} \beta_s\left(\frac{|\lambda|_\infty}{\mu}\right).$$

(5.10)
6 Example: Changing $M/M/s/0$ Rates in Midstream

Suppose we consider the case of $\lambda(t) = \lambda_+$ and $\mu(t) = \mu_+$ for all $t \geq 0$ and $p(0) = p^*(0)$ where $m^*(0) = \lambda_-/\mu_-$. The time-dependent behavior of $Q$ is that of a stationary $M/M/s/0$ queue with rates $\lambda_-$ and $\mu_-$ for all time $t < 0$, that suddenly switches to rates $\lambda_+$ and $\mu_+$ for all time $t \geq 0$. We now want to compute an upper bound for the error between the transient distribution of $Q$ and its MOL approximation. Now, in addition to Theorem 4.1, we exploit the fact that $E_A(t) = \exp(At)$, where $A$ is the infinitesimal generator of the $M/M/s/0$ queue with parameters $\lambda_+$ and $\mu_+$. In this case, (4.9) becomes

$$p^*(t) - p(t) = \int_0^t \beta(m^*(\tau), s) \cdot (p^*(\tau) - e_s) \exp((t - \tau)A) dm^*(\tau).$$  \hspace{1cm} (6.1)

If we let $\rho_+ = \lambda_+ / \mu_+$ and $\rho_- = \lambda_- / \mu_-$, then $m^*(\tau) = \rho_+ \exp(-\mu_+ \tau) + \rho_+(1 - \exp(-\mu_+ \tau))$ and

$$d \frac{dm^*(\tau)}{d\tau} = \mu_+(\rho_+ - \rho_-) \exp(-\mu_+ \tau).$$  \hspace{1cm} (6.2)

Since $Q$ is reversible, see page 32 of Keilson [9], the generator $A$ is diagonally similar to a symmetric matrix. By the spectral decomposition theorem, we have

$$\exp(tA) = 1^T \pi + \sum_{j=1}^s \exp(-\sigma_j t) \Delta(\sqrt{\pi})^{-1} x_j^T x_j \Delta(\sqrt{\pi}),$$  \hspace{1cm} (6.3)

where $\pi$ is the steady-state probability vector for $A$ (which satisfies $\pi A = 0$), $\sqrt{\pi}$ is the positive vector whose components are the square roots of the components of $\pi$, and $\Delta(\sqrt{\pi})$ is the corresponding diagonal matrix. The negatives of the $s + 1$ real numbers $0 < \sigma_1 < \cdots < \sigma_s$ are the eigenvalues for $A$. Finally, $\{\sqrt{\pi}, x_1, \ldots, x_s\}$ is the corresponding set of orthonormal eigenvectors for $\Delta(\sqrt{\pi}) A \Delta(\sqrt{\pi})^{-1}$. The eigenvalues and eigenvectors for this model are readily obtained, as shown by Lederman and Reuter [11] and Karlin and McGregor [8]. In this case the orthogonal polynomials are the Poisson-Charlier polynomials; also see Jagerman [7]. This leads to the following inequality:

$$|p^*(t) - p(t)| \leq 2 \mu_+ b(\tilde{\rho}, s) |\rho_+ - \rho_-| \cdot \sqrt{\frac{1}{\pi_k} \sum_{j=1}^s \frac{\exp(-\sigma_j t) - \exp(-\mu_+ t)}{\mu_+ - \sigma_j}},$$  \hspace{1cm} (6.4)

where $\tilde{\rho} = \max(\rho_+, \rho_-)$.

In addition to these error estimates, we get the following stochastic dominance results by applying Corollary 4.3,

$$\rho_- \leq \rho_+ \implies p^*(t) \preceq_{st} p(t) \text{ for all } t \geq 0,$$  \hspace{1cm} (6.5)

and

$$\rho_- \geq \rho_+ \implies p^*(t) \succeq_{st} p(t) \text{ for all } t \geq 0.$$  \hspace{1cm} (6.6)
References


