Operations Research Letters 42 (2014) 293–298

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

A Poisson–Charlier approximation for nonstationary queues

Jamol Pender

School of Operations Research and Information Engineering, Cornell University, 206 Rhodes Hall, Ithaca, NY 14850, USA

ARTICLE INFO

Article history:
Received 24 February 2014
Received in revised form
28 March 2014
Accepted 3 May 2014
Available online 17 May 2014

Keywords:
Multi-server queues
Abandonment
Time-varying rates
Poisson–Charlier polynomials

ABSTRACT

In this paper, we develop a new approximation for nonstationary multiserver queues with abandonment. Our method uses the Poisson–Charlier polynomials, which are a discrete orthogonal polynomial sequence that is orthogonal with respect to the Poisson distribution. We show that by appealing to the Poisson–Charlier polynomials that we can estimate the mean, variance, and probability of delay of our nonstationary queueing system with good accuracy. Lastly, we provide a numerical example that illustrates that our approximations are effective.

1. Introduction

Many real time service processes can be modeled using nonstationary queueing models. Some of the most prevalent applications of nonstationary queueing models include but are not limited to telecommunication networks, healthcare systems, call centers, hospitality networks, airline reservations, and transportation systems. See for example the papers of [1,3,4]. Healthcare systems in particular often are subject to nonstationary dynamics that depends on the time of day and the state of the system. In fact, mass casualty events and large scale accidents, changes in demand, and the availability of nurses and beds are just some of the many ways that healthcare systems can experience transient and nonstationary dynamics.

Fluid and diffusion limit theorems are popular tools to analyze nonstationary queueing systems. Indeed authors of [8] proved fluid and diffusion limit theorems for these nonstationary systems when the processes are Markovian and [7] proved fluid and diffusion limit theorems when the service and abandonment distributions are non-exponential. However, as pointed out by [6], these limit theorems are not valid when the arrival rate or the number of servers is not large. Thus, new methods must be developed in order to estimate the nonstationary dynamics of queues with abandonment.

As the model that we consider is Markovian, we exploit the functional Kolmogorov forward equations for the queueing process. Like [16] or [15], we encounter difficulties from the forward equations since the differential equations are not a *closed* system. This definition of *closed* should not be confused with a closed queueing process with no external arrivals. When a dynamical system is not closed, this implies that the forward equations for the nth moment cannot be written in terms of functions of the nth moment and lower order moments, see for example [13] for a mathematical definition. However, there are three exceptions to when the dynamics of the multiserver queue with abandonment is a closed dynamical system. They are the infinite server, a no-server queue, or when the mean service rate is equal to the mean abandonment rate. Thus, it is necessary to prescribe an appropriate surrogate distribution for the queue length process in order to close the forward equations, which is non-trivial for systems with abandonment.

Our new approach to approximate the dynamics of the queueing process is to expand the queue length process in terms of a truncated sequence of Poisson–Charlier polynomials. This method is unlike the methods of [9,10,14,13] since we use a discrete orthogonal polynomial sequence, which is more natural for approximating the discrete queueing process. Moreover, the Poisson–Charlier polynomials are orthogonal with respect to the Poisson distribution. Using a polynomial sequence that is related to the Poisson distribution is also natural because in the nonstationary setting when the queueing process is an infinite server queue or a birth–death process with a deterministic birth rate and a linear death rate, the distribution is a Poisson distribution that changes dynamically through time. Thus, by using the Poisson–Charlier polynomials, we preserve the discrete nature of the queueing process, while being able to provide accurate estimates for various performance measures.

E-mail address: jjp274@cornell.edu.

http://dx.doi.org/10.1016/j.orl.2014.05.001
0167-6377/

2. The nonstationary queueing model

In this section we first consider the nonstationary Erlang-A queue, which is known as the $M_t/M_t/c_t + M_t$ queue. In the work of Mandelbaum et al. [8], they show that the Erlang-A queueing process $Q \equiv \{Q(t)|t \geq 0\}$ can be represented by the following stochastic, time changed integral equation:

$$Q(t) = Q(0) + \Pi_1\left(\int_0^t \lambda(s)ds\right) - \Pi_2\left(\int_0^t \mu \cdot (Q(s) \wedge c(s))ds\right) - \Pi_3\left(\int_0^t \beta \cdot (Q(s) - c(s))^+ ds\right),$$

where $\Pi_i \equiv \{\Pi_i(t)|t \geq 0\}$ for $i = 1, 2, 3$ are i.i.d. standard (rate 1) Poisson processes. Deterministic and random time changes easily make the three unit rate Poisson processes into counting processes that keep track of the arrivals, service departures, and abandonments respectively.

The Erlang-A model is not only an important model to study because of its ability to approximate many service systems, but also it is a special case of a *Markovian service network*. Like in the work of [8], this means that we can construct an associated, scaled or *uniformly accelerated* queueing process where the new arrival rate function is $\eta \cdot \lambda$ and the new number of servers is $\eta \cdot c$ for some scale factor $\eta > 0$. Taking the following pointwise limits gives us the *fluid* models of [8], where convergence is with respect to the space $D[0, \infty)$ with the J_1 topology, i.e.

$$\lim_{\eta\to\infty} \frac{Q^\eta}{\eta} = q \quad \text{a.s. u.o.c}$$

where the deterministic process q, the *fluid mean*, is governed by the following one dimensional dynamical system:

$$\overset{\bullet}{q} = \lambda - \mu \cdot (q \wedge c) - \beta \cdot (q - c)^+. \quad (1)$$

If we subtract the fluid limit and rescale we obtain the diffusion limits of [8] i.e.

$$\lim_{\eta\to\infty} \sqrt{\eta} \cdot \left(\frac{Q^\eta}{\eta} - q\right) \overset{d}{=} \hat{Q}.$$

Moreover, as pointed out in [8], if the set of time points $\mathcal{A}$

$$\mathcal{A} \equiv \{t \mid q(t) = c(t)\}$$

has measure zero, then $\hat{Q}$ is a Gaussian diffusion process whose variance combines with the fluid mean to form the following two-dimensional dynamical system:

$$\overset{\bullet}{q} = \lambda - \mu \cdot (q \wedge c) - \beta \cdot (q - c)^+$$
$$\overset{\bullet}{v} = \lambda + \mu \cdot (q \wedge c) + \beta \cdot (q - c)^+ - 2 \cdot \mu \cdot v \cdot \{q < c\} - 2 \cdot \beta \cdot v \cdot \{q > c\}$$

where $\{q < c\}$ denotes an *indicator function* that equals one if the statement is true i.e. if $q < c$, and zero if the statement is false.

As noted in [9], the fluid and diffusion limits provide good estimates of the performance of the queueing process when the rates are large and the set $\mathcal{A}$ has measure zero. However, there are many practical reasons why the set $\mathcal{A}$ should not have measure zero. From a fluid optimal control perspective, often the optimal staffing level is when the number of servers is equal to the fluid limit, see for example [3]. Moreover, in some service settings such as a moderate sized hospital, it is not clear that the rates should be very large either. This inaccuracy of the fluid and diffusion limits, motivated [9] to explore alternative methods that can approximate the performance of the queueing model in a large or small rate regime.

2.1. Functional Kolmogorov forward equations

Our approach to studying the Erlang-A model is to use the functional Kolmogorov forward equations. The functional version of the forward equations for the $M_t/M_t/c_t + M_t$ queue has the following form:

$$\overset{\bullet}{E}[f(Q)] = \lambda(t) \cdot E[f(Q+1) - f(Q)] + \mu(t) \cdot E[(Q \wedge c) \cdot (f(Q-1) - f(Q))] + \beta(t) \cdot E\left[(Q - c)^+ \cdot (f(Q-1) - f(Q))\right],$$

for all appropriate functions f whose expectations are well defined. For notational convenience we suppress the time dependence of λ, μ, β, c. Moreover, we use the "•" notation to denote its time derivative i.e.

$$\overset{\bullet}{E}[f(Q)] \equiv \frac{d}{dt} E[f(Q(t))].$$

For special cases of f such as Q, $(Q - E[Q])^2$, we obtain the following set of cumulant moment, Kolmogorov forward equations for the mean and variance as:

$$\overset{\bullet}{E}[Q] = \lambda - \mu \cdot E[Q \wedge c] - \beta \cdot E\left[(Q - c)^+\right]$$
$$\overset{\bullet}{\text{Var}}[Q] = \lambda + \mu \cdot E[Q \wedge c] + \beta \cdot E\left[(Q - c)^+\right] - 2\left(\mu \cdot \text{Cov}[Q, Q \wedge c] + \beta \cdot \text{Cov}\left[Q, (Q - c)^+\right]\right).$$

Although it appears that we have exact formulas for the mean and variance of the forward equations, one must observe that in order to compute the expectation and covariance terms, we need to know the true distribution of the queue length process. However, we do not know the distribution of the queue length process a priori. This unknown distribution motivated [9,10,14,13] to approximate the queueing process by a surrogate distribution. However, their approach was motivated by fluid and diffusion limit theorems and use continuous surrogate distributions that have mass on the entire real line and do not capture the discrete nature of the queueing process.

2.2. Ineffectiveness of the Poisson surrogate distribution

The first discrete distribution that arises when considering multiserver queues is the Poisson distribution. The Poisson distribution is quite natural since it is the distribution of an infinite server queue, even in the nonstationary setting. In [2], they use the properties of the Poisson arrival process and exploit Poisson random measures to show that the $M_t/G/\infty$ queue $Q^\infty(t)$, has a Poisson distribution with time varying rate $q^\infty(t)$. The exact analysis of the infinite server queue is often useful since it represents the dynamics of the queueing process if there was an unlimited amount of resources to satisfy the demand process. As observed in [2], the mean of the queue length process $q^\infty(t)$ has the following representation:

$$q^\infty(t) \equiv E[Q^\infty(t)] = \int_{-\infty}^t \overline{G}(t-u)\lambda(u)du = E\left[\int_{t-S}^t \lambda(u)du\right] = E[\lambda(t - S_e)] \cdot E[S]$$

where S represents a service time with distribution G, $\overline{G} = 1 - G(t) = \mathbb{P}(S > t)$, and S_e is a random variable with distribution that follows the stationary excess of residual-lifetime cdf G_e, defined by

$$G_e(t) \equiv \mathbb{P}(S_e < t) = \frac{1}{E[S]} \int_0^t \overline{G}(u)du, \quad t \geq 0.$$

Thus, our first idea is to use a Poisson distribution to approximate the dynamics of the queueing process. This is equivalent to projecting our queueing process onto a birth–death process with

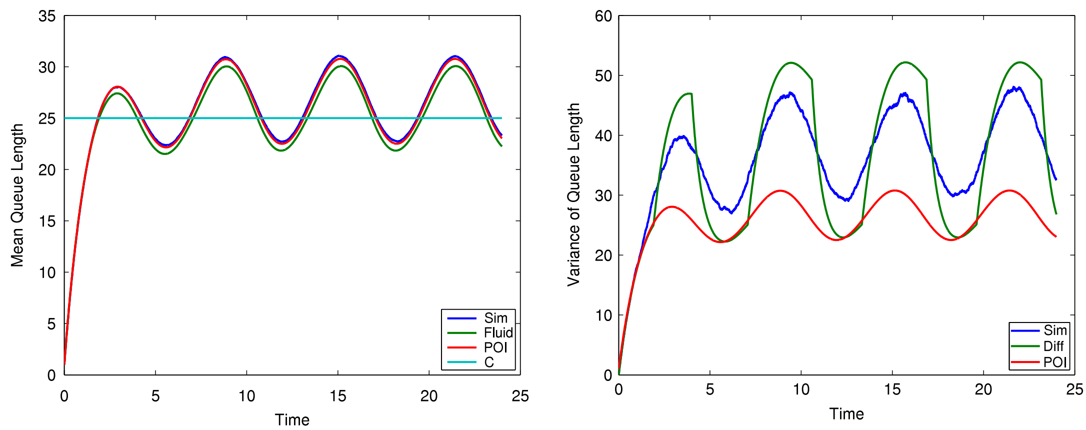

Fig. 1. Mean (Left). Variance (Right). $\lambda = 25 + 5 \cdot \sin(t)$, $\mu = 1$, $\beta = 0.5$, $T = 24$, $q(0) = 1$.

a nonstationary and deterministic arrival rate and a linear death rate. This is equivalent to

$$Q(t) \overset{D}{\approx} \text{Poisson}(q(t)) \tag{2}$$

for all $t \geq 0$, where $\{q(t)|t \geq 0\}$ is some one-dimensional dynamical system. This Poisson approximation for the queue length process leads us to our first theorem.

Theorem 2.1. *The rate functions for the mean of the functional forward equations have the following closed form expressions when queue length distribution is given by Eq.* (2)

$$E[(Q \wedge c)] = q - q \cdot \Gamma(c-1, q) + c \cdot \Gamma(c, q) \tag{3}$$

$$E\left[(Q - c)^+\right] = q \cdot \Gamma(c-1, q) - c \cdot \Gamma(c, q). \tag{4}$$

Remark 2.2. It is important to mention that we do not consider the functional forward equations for the variance when we assume that the queue length distribution is Poisson. This is because the Poisson distribution is characterized by its mean and all cumulant moments of the Poisson are equal to its mean. Thus, the variance is identical to the mean.

Proof. To prove this result, we must show that under the Poisson $(q(t))$ assumption, that the rate functions of the mean satisfy the closed form expressions of Eqs. (3) and (4). Before we prove the results, we provide a lemma that shows that the tail distribution of a Poisson distribution can be expressed in terms of the incomplete gamma function.

Lemma 2.3.

$$\Gamma(c, x) = \sum_{m=c+1}^{\infty} e^{-x} \cdot \frac{x^m}{m!} = \frac{1}{\Gamma(c)} \int_0^x e^{-y} y^{c-1} dy$$

$$\overline{\Gamma}(c, x) = \sum_{m=0}^{c} e^{-x} \cdot \frac{x^m}{m!} = \frac{1}{\Gamma(c)} \int_x^{\infty} e^{-y} y^{c-1} dy.$$

Proof. See [5]. □

Now to complete the proof of the theorem, we have for the service rate function that

$$\begin{aligned} E[(Q \wedge c)] &= E[Q] - E\left[(Q-c)^+\right] \\ &= q - E\left[(Q-c) \cdot \{Q \geq c\}\right] \\ &= q - \sum_{m=c+1}^{\infty} m \cdot e^{-q} \cdot \frac{q^m}{m!} + c \cdot \sum_{m=c+1}^{\infty} e^{-q} \cdot \frac{q^m}{m!} \\ &= q - q \cdot \sum_{m=c}^{\infty} e^{-q} \cdot \frac{q^m}{m!} + c \cdot \Gamma(c, q(t)) \\ &= q \cdot \overline{\Gamma}(c-1, q(t)) + c \cdot \Gamma(c, q(t)). \end{aligned}$$

Moreover, for the abandonment rate function, we have that

$$\begin{aligned} E\left[(Q-c)^+\right] &= E[Q] - E\left[(Q \wedge c)\right] \\ &= q - q \cdot \overline{\Gamma}(c-1, q(t)) - c \cdot \Gamma(c, q(t)) \\ &= q \cdot \Gamma(c-1, q(t)) - c \cdot \Gamma(c, q(t)). \quad \square \end{aligned}$$

We see on the left of Fig. 1 that the fluid limit is doing a good job of estimating the mean performance. However, on the right of Fig. 1 we see that the diffusion limit is not doing well at estimating the performance of the queueing system. Moreover, we see that the diffusion limit is the most inaccurate when the fluid limit is near the number of servers i.e. $q(t) \approx c(t)$. We also see that the Poisson approximation is also doing well for the mean on the left of Fig. 1, but fails for the variance on the right of Fig. 1. Thus, we are led to find a better approximation for the queueing process that is also centered around the Poisson process and this leads to the Poisson–Charlier polynomials.

3. The Poisson–Charlier approximation

In this section, we describe how to use Poisson–Charlier polynomials in conjunction with the functional forward equations in order to construct approximations for our nonstationary queueing processes. The Poisson–Charlier polynomials are an orthogonal polynomial sequence with respect to the Poisson distribution with rate α i.e.

$$\phi(\alpha, k) = e^{-\alpha} \frac{\alpha^k}{k!} \quad k = 0, 1, 2, \ldots.$$

As a result, the Poisson–Charlier polynomials solve the following recurrence relation:

$$C_{n+1}(k, \alpha) = (k - n - \alpha) \cdot C_n(k, \alpha) - n \cdot \alpha \cdot C_{n-1}(k, \alpha).$$

The first three Poisson–Charlier polynomials are defined as

$$\begin{aligned} C_0(k, \alpha) &= 1 \\ C_1(k, \alpha) &= k - \alpha \\ C_2(k, \alpha) &= k^2 - 2 \cdot k \cdot \alpha + \alpha^2 - k. \end{aligned}$$

Now suppose that we have a function $f(k)$, which is defined on the integers and satisfies the inequality

$$\sum_{k=0}^{\infty} f^2(k) \phi(k, \alpha) < \infty, \quad \text{for some } \alpha > 0.$$

Then we have the following expansion in terms of Poisson–Charlier polynomials in the Hilbert space $l^2(\mathbb{N}, \phi(\alpha, k))$.

Proposition 3.1. *Any function $f(k) \in l^2(\mathbb{N}, \phi(\alpha, k))$ can be expanded into a Poisson–Charlier series i.e.*

$$f(k) = \sum_{m=0}^{\infty} q_m \cdot C_m(k, \alpha) \tag{5}$$

where $q_m = \sum_{k=0}^{\infty} f(k) C_m(k, \alpha)\phi(k, \alpha)$.

Proof. See [11]. □

Remark 3.2. This expansion can also be extended to the case where the independent variable of the function $f(k)$ is a stochastic process and also depends on time itself.

Inspired by the expansion of the function $f(k)$ on the integers, we can apply this expansion to our queueing process $Q(t)$ where

$$Q(t) \stackrel{d}{=} \sum_{m=0}^{\infty} q_m(t) \cdot C_m(k, \alpha). \tag{6}$$

Moreover, if we only use the first n polynomials to approximate our queueing process, we have that

$$Q^{(n)}(t) \stackrel{d}{=} \sum_{m=0}^{n} q_m(t) \cdot C_m(k, \alpha). \tag{7}$$

3.1. First order expansion

The first order expansion of our method uses the first polynomial of the Poisson–Charlier family. The first polynomials is the identity and yields a deterministic constant i.e.

$$Q^{(0)}(t) \stackrel{d}{=} q_0(t) \equiv q(t).$$

Using the deterministic function for as an approximation for the queue length leads us to our first theorem.

Theorem 3.3. *By substituting $q(t)$ into the functional Kolmogorov forward equations for the mean of the queue length process, we have that*

$$\overset{\bullet}{q} = \lambda - \mu \cdot (q \wedge c) - \beta \cdot (q - c)^+.$$

Proof.

$$\begin{aligned}
\overset{\bullet}{E}[Q] &= \overset{\bullet}{q} \\
&= \lambda - \mu \cdot E[Q \wedge c] - \beta \cdot E\left[(Q - c)^+\right] \\
&= \lambda - \mu \cdot E[q \wedge c] - \beta \cdot E\left[(q - c)^+\right] \\
&= \lambda - \mu \cdot (q \wedge c) - \beta \cdot (q - c)^+. \quad \square
\end{aligned}$$

The resulting process is the same as the fluid limit of the queueing process of [8] and the deterministic mean approximation (DMA) of [10]. It was shown in [10] that the DMA estimates the mean of the queue length quite well when the number of servers is large and the queue is not critically loaded. However, the DMA does not estimate the mean well when the number of servers is small or when the queue is critically loaded. Furthermore, DMA gives no insight into the variations of the queueing process since it implicitly assumes that these variations are zero. As a result, we are led to another approximation that is not purely deterministic and incorporates the stochastic behavior of the queueing process into the estimation of the functional forward equations.

3.2. Second order Poisson–Charlier expansion

In this section, we add an additional Poisson–Charlier polynomial to approximate the queue length distribution. This second approximation developed is called the *Second Order Poisson–Charlier Expansion* (SOPCE). The SOPCE is constructed by assuming that $Q \equiv \{Q(t) | t \geq 0\}$ such that

$$Q(t) \stackrel{d}{=} q(t) + \sqrt{v(t)} \cdot \frac{K - \alpha}{\sqrt{\alpha}} \tag{8}$$

for all $t \geq 0$, where $\{q(t), v(t) | t \geq 0\}$ is some two-dimensional, deterministic, dynamical system where the v process is always positive and K is a Poisson random variable with rate α. This new approximation for the queue length distribution is based on the Poisson distribution, however, it allows for a mean and variance that are not identical. This approximation also leads us to our next theorem using the functional forward equations.

Theorem 3.4. *The rate functions for the mean and variance of the functional forward equations have the following closed form expressions when queue length distribution is given by Eq. (8)*

$$\begin{aligned}
&E\left[(Q - c)^+\right] = \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot (\alpha \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) - \xi \cdot \Gamma(\lceil \xi \rceil, \alpha)) \\
&E[Q \wedge c] = q - \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot (\alpha \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) - \xi \cdot \Gamma(\lceil \xi \rceil, \alpha)) \\
&\mathrm{Cov}\left[Q, (Q - c)^+\right] = v \cdot \alpha \cdot \Gamma(\lceil \xi \rceil - 2, \alpha) + v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) \\
&\qquad - (\alpha + \xi) \cdot v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) + v \cdot \xi \cdot \Gamma(\lceil \xi \rceil, \alpha) \\
&\mathrm{Cov}[Q, Q \wedge c] = v - v \cdot \alpha \cdot \Gamma(\lceil \xi \rceil - 2, \alpha) - v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) \\
&\qquad + (\alpha + \xi) \cdot v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) - v \cdot \xi \cdot \Gamma(\lceil \xi \rceil, \alpha), \\
&\xi = \alpha + \sqrt{\alpha} \cdot \frac{c - q}{\sqrt{v}}.
\end{aligned}$$

Proof. It suffices to show the proof for the terms of the max function $(Q - c)^+$ since we have that $(Q - c)^+ = Q - Q \wedge c$.

$$\begin{aligned}
E\left[(Q - c)^+\right] &= E\left[\left(q + \sqrt{v} \cdot \frac{K - \alpha}{\sqrt{\alpha}} - c\right)^+\right] \\
&= \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot E\left[(K - \xi)^+\right] \\
&= \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} (m - \xi) \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&= \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} m \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&\quad - \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} \xi \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&= \frac{\sqrt{v}}{\sqrt{\alpha}} \cdot (\alpha \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) - \xi \cdot \Gamma(\lceil \xi \rceil, \alpha))
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Cov}\left[Q, (Q - c)^+\right] = E\left[Q \cdot (Q - c)^+\right] - E[Q] \cdot E\left[(Q - c)^+\right] \\
&\quad = \frac{v}{\alpha} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} (m - \alpha) \cdot (m - \xi) \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&\quad = \frac{v}{\alpha} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} (m^2 - (\alpha + \xi) \cdot m + \alpha \cdot \xi) \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&\quad = \frac{v}{\alpha} \cdot \sum_{m=\lceil \xi \rceil}^{\infty} (m^2 - (\alpha + \xi) \cdot m + \alpha \cdot \xi) \cdot e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\
&\quad = v \cdot \alpha \cdot \Gamma(\lceil \xi \rceil - 2, \alpha) + v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) \\
&\qquad - (\alpha + \xi) \cdot v \cdot \Gamma(\lceil \xi \rceil - 1, \alpha) + v \cdot \xi \cdot \Gamma(\lceil \xi \rceil, \alpha). \quad \square
\end{aligned}$$

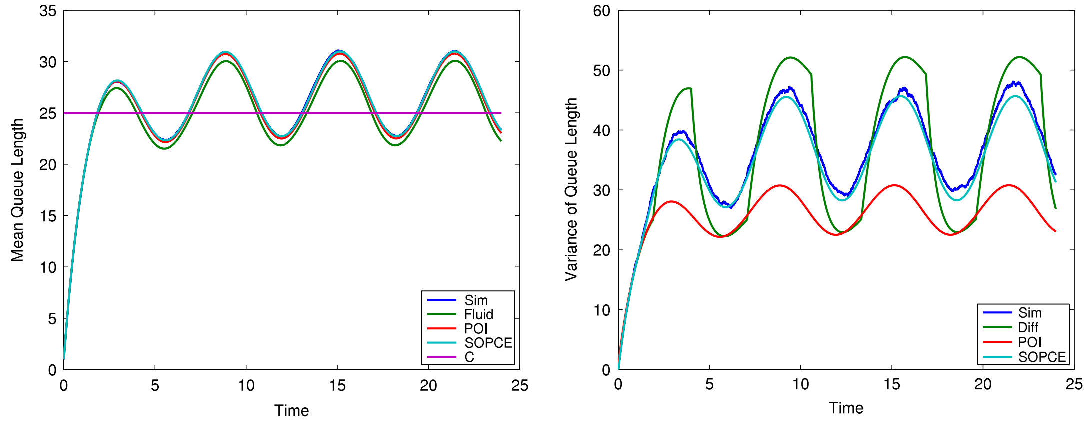

Fig. 2. Mean (Left). Variance (Right). $\lambda = 25 + 5 \cdot \sin(t)$, $\mu = 1$, $\beta = 0.5$, $T = 24$, $q(0) = 1$.

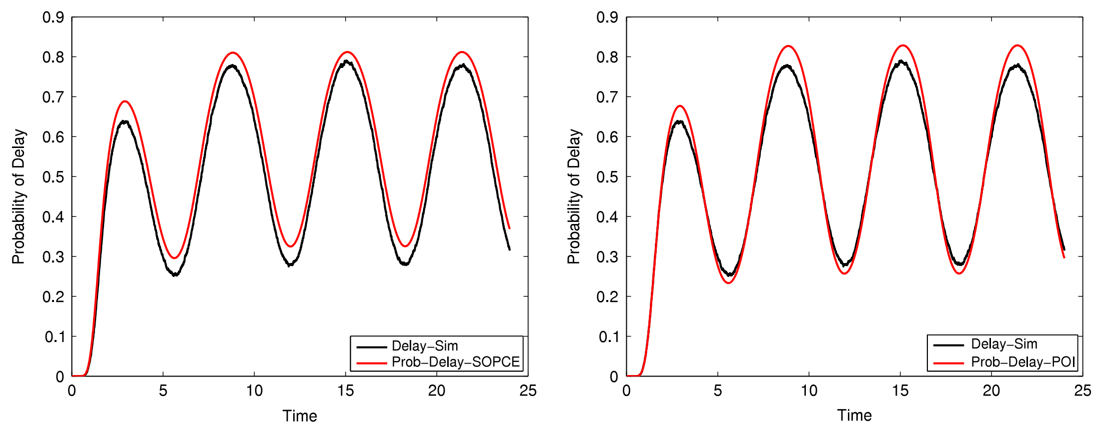

Fig. 3. Poisson delay probability (Left). Poisson–Charlier delay probability (Right). $\lambda = 25 + 5 \cdot \sin(t)$, $\mu = 1$, $\beta = 0.5$, $T = 24$, $q(0) = 1$.

On the left of Fig. 2 we see that the second order Poisson–Charlier approximation is also approximating the mean of the queueing process very well. Moreover, on the right of the Fig. 2 we see that the SOPCE is also estimating the variance of the queueing process accurately. This is because we are able to separate the mean and variance of the distribution unlike the Poisson distribution, which has the same distribution and is defined only through its mean. Thus, the SOPCE is the approximation that is working quite well.

3.3. Probability of delay

In this section, we derive a simple approximation for the delay probability of the queueing process using the SOPCE method. The delay probability can be approximated by

$$\begin{aligned}\mathbb{P}(Q \geq c) &= \mathbb{P}(K \geq \xi) \\ &= \sum_{m=\lceil \xi \rceil}^{\infty} e^{-\alpha} \cdot \frac{\alpha^m}{m!} \\ &= \Gamma(\lceil \xi \rceil, \alpha).\end{aligned}$$

On the left of Fig. 3 we see that the Poisson approximation does a decent job of approximating the delay probability, however, on the right of Fig. 3 we see that the SOPCE does a better job of estimating the delay probability of the queueing process. This is also because we are able to separate the mean and variance of the distribution unlike the Poisson distribution, which has the same distribution and is defined only through its mean. Thus, the SOPCE performs better than the Poisson surrogate distribution.

4. Conclusion and contributions

We have shown that the Poisson–Charlier polynomials can be used to generate approximations for the dynamics of nonstationary queueing systems. We have illustrated that our method is able to estimate the mean, variance, and probability of delay with good accuracy. Although we have applied this theory to nonstationary queueing processes, our method also generally applies to any Markovian birth–death process such as nonstationary loss queues where the functional forward equations have the following representation like in [12]

$$\begin{aligned}\overset{\bullet}{E}[f(Q)] &= \lambda(t) \cdot E[(f(Q+1) - f(Q)) \cdot \{Q < c + k\}] \\ &\quad + \mu(t) \cdot E[(Q \wedge c) \cdot (f(Q-1) - f(Q))] \\ &\quad + \beta(t) \cdot E\left[(Q-c)^+ \cdot (f(Q-1) - f(Q))\right].\end{aligned}$$

Moreover, our method can be applied to nonstationary stochastic epidemic systems where the forward equations are also not a closed system. Lastly, the use of the Poisson distribution is not unique. One can also apply the same method using the negative binomial distribution where the Meixner polynomials are especially relevant. We hope to explore the Meixner polynomials in the infinite support setting and the Hahn and Racah polynomials for birth–death processes with finite support. Lastly, we believe that we can continue the expansion method to study other cumulant moments such as the skewness and kurtosis, however, this can easily be continued via the method of [10].

References

[1] I. Cohen, A. Mandelbaum, N. Zychlinski, Minimizing mortality in a mass casualty event: fluid networks in support of modeling and staffing, IIE Trans. (2013) in press.

[2] S.G. Eick, W.A. Massey, W. Whitt, The Physics of the $M_t/G/\infty$ Queue, Oper. Res. 41 (1993) 731–742.

[3] R. Hampshire, O.B. Jennings, W.A. Massey, A time varying call center design with Lagrangian mechanics, Probab. Engrg. Inform. Sci. 23 (2009) 231–259.

[4] R. Hampshire, W.A. Massey, A tutorial on dynamic optimization and applications to queueing systems with time-varying rates, Tutor. Oper. Res. (2010) 208–247.

[5] A.J.E.M. Janssen, J.S. van Leeuwaarden, B. Zwart, Gaussian expansions and bounds for the Poisson distribution applied to the Erlang B formula, Adv. Appl. Probab. 40 (2008) 122–143.

[6] Y.M. Ko, N. Gautam, Critically loaded time-varying multiserver queues: computational challenges and approximations, INFORMS J. Comput. 25 (2013) 285–301.

[7] Y. Liu, W. Whitt, Many-server heavy-traffic limits for queues with time-varying parameters, Ann. Appl. Probab. 24 (2014) 378–421.

[8] A. Mandelbaum, W.A. Massey, M. Reiman, Strong approximations for Markovian service networks, Queueing Syst. 30 (1998) 149–201.

[9] W.A. Massey, J. Pender, Skewness variance approximation for dynamic rate mutliserver queues with abandonment, Perform. Eval. Rev. 39 (2011) 74–74.

[10] W.A. Massey, J. Pender, Gaussian Skewness approximation for dynamic rate multiserver queues with abandonment, Queueing Syst. 75 (2013) 243–277.

[11] H. Ogura, Orthogonal functionals of the Poisson process, IEEE Trans. Inform. Theory 18 (1972) 473–481.

[12] J. Pender, Cumulant Moment Approximations for Nonstationary Loss Queues, Cornell University, 2013, Available at: http://www.columbia.edu/~jp3404.

[13] J. Pender, Laguerre Polynomial Approximations for Nonstationary Queues, Cornell University, 2013, Available at: http://www.columbia.edu/~jp3404.

[14] J. Pender, Gram Charlier expansions for time varying multiserver queues with abandonment, SIAM J. Appl. Math. (2014) in press.

[15] M.H. Rothkopf, S.S. Oren, A closure approximation for the nonstationary $M/M/s$ queue, Manag. Sci. 25 (1979) 522–534.

[16] M.R. Taaffe, K.L. Ong, Approximating nonstationary $\mathrm{Ph}(t)/M(t)/s/c$ queueing systems, Ann. Oper. Res. 8 (1987) 103–116.