Probability in the Engineering and Informational Sciences, **29**, 2015, 27–49.
doi:10.1017/S0269964814000205

NONSTATIONARY LOSS QUEUES VIA CUMULANT MOMENT APPROXIMATIONS

JAMOL PENDER
School of Operations Research and Information Engineering
Cornell University, Ithaca, NY 14850, USA
E-mail: jjp274@cornell.edu

In this paper, we provide a new technique for analyzing the nonstationary Erlang loss queueing model with abandonment. Our method uniquely combines the use of the functional Kolmogorov forward equations with the well-known Gram-Charlier series expansion from the statistics literature. Using the Gram-Charlier series expansion, we show that we can estimate salient performance measures of the loss queue such as the mean, variance, skewness, kurtosis, and blocking probability. Lastly, we provide numerical examples to illustrate the effectiveness of our approximations.

1. INTRODUCTION

Many real-time service processes can be modeled using nonstationary Erlang loss queueing models. Some applications of nonstationary loss queues include but are not limited to telecommunication networks, healthcare systems, call centers, hospitality networks, airline reservations, and transportation systems. See, for example, Grier, Massey, McKoy and Whitt [2], Hampshire et al. [3–5]. Communication networks in particular often are subject to a multitude of nonstationary dynamics that depend on the time of day and the state of the system. In fact, buffer overflows, changes in demand, and the availability of service are just some of the many ways that communication systems can experience transient and nonstationary dynamics. Moreover, when the arrival process explicitly depends on the time of day, nonstationary models are inevitable.

The stationary Erlang loss model, which we denote by $M/M/c/0$, has a Poisson arrival process, independent and identically distributed service times from an exponential distribution, and c parallel servers with no extra waiting spaces. What makes the Erlang loss system different from the standard multiserver queue is that if a customer arrives while the all the servers are busy, then that customer is immediately lost and never receives service. Although most communication networks experience nonstationary conditions, much of the literature only considers stationary processes. Moreover, much of the literature for stationary processes, does not carry over quite easily to nonstationary models, and requires more insight and analysis.

Much of the research on the nonstationary Erlang loss model has focused on approximating the *blocking probability*, which is perhaps the most important performance measure of the Erlang loss model. One such approximation method for estimating the blocking

 0269-9648/15 $25.00

probability is the well-known *modified offered load approximation* of Jagerman [6]. It uses the mean offered load from an infinite server queue and naively substitutes it into the Erlang blocking formula as the mean offered load. Massey and Whitt [11], rigorously show that the modified offered load approximation is appropriate when the blocking probability is small and the loss queue is well approximated by the infinite server queue. Furthermore, they also provide bounds for the performance of the modified offered load approximation based on the input parameters, which is quite useful in practice. In another paper, Massey and Whitt [12] show how to use a non-Poisson, but stationary arrival process with a higher coefficient of variation to approximate the blocking probability induced by a time-varying arrival rate. Lastly, in the paper of Davis, Massey and Whitt [1], they show that the nonstationary loss queue blocking probability is not insensitive to the service distribution and depends significantly on the variance of the service distribution.

In addition to nonstationary arrivals, the traditional Erlang loss model does not capture the realistic phenomenon known as abandonment. It is well-known that customers do not have infinite patience and are likely to renege from the system if the time that they must wait for service is deemed to be excessive. Thus, in our model, we also add customer abandonment if customers are forced to spend time in the available waiting spaces of the queue. Without the features of time-varying arrivals and abandonment, our model is exactly the $M/M/c/k$ queueing model, which was studied extensively in the dissertation of Wallace [16] where rigorous asymptotics of the $M/M/c/k$ queue were developed. The dissertation of [16] also derives many closed-form expressions for blocking and delay probabilities. However, the nonstationary dynamics precludes us from deriving closed form expressions for the queueing behavior.

Besides the nonstationary arrivals and the inclusion of abandonment, our model is quite different from a traditional multiserver queueing model in that the arrival process is actually *state-dependent*. Moreover, the state dependence is discontinuous with respect to the queue length process. Thus, we cannot leverage the fluid and diffusions approximations for Markovian service networks of Mandelbaum, Massey and Reiman [8]. One way around this is to incorporate what is known as *fast abandonment* like in the work of Hampshire et al. [4]. However, if the fast abandonment parameter is not large enough, one could have occasional situations where the queue length exceeds the threshold, which is not allowed and biases the queue length to larger values than expected from the standard loss queue. Thus, it is imperative that we develop new techniques for analyzing the dynamic behavior of the nonstationary loss queue with abandonment.

In this paper, we propose using the exact stochastic process via the functional forward equations and combining it with Gram-Charlier series expansions from the statistics literature. We should mention that we are not the first to consider using the functional forward equations to approximate the time-dependent moments of queueing process. Authors such as Massey and Pender [10] use the functional forward equations with a novel expansion of the queue length process in terms of Hermite polynomials for multi-server queues with abandonment. However, a major difference is that [10] expands the queue length process while we expand the density. Moreover, Pender [13] uses the Gram-Charlier series approach, however, only applied it to the multiserver queue, which also fits within the Markovian service network family. However, we are the first to apply the Gram-Charlier series approach to queueing processes that do not fit into the Markovian service network framework, and also are the first to study the time-dependent mean, variance, skewness, kurtosis, and blocking probability of nonstationary loss queues with abandonment. Moreover, our approximations for the blocking probability are accurate even when the blocking probability is not small. This is a significant advance in approximating the loss queue since many approximations

assume that the loss queue blocking probabilities are small and thus the loss queue is well approximated by an infinite server queue.

1.1. Contributions

To the best of our knowledge our contributions in this work are the following.

- We combine the functional Kolmogorov forward equations with Gram-Charlier series expansions to develop novel approximations for nonstationary loss queues.
- We derive accurate approximations for the mean, variance, skewness, and kurtosis of the nonstationary loss queue with abandonment.
- We illustrate that higher-order moments of the nonstationary loss queue can improve the estimates of the lower moments.
- We avoid the use of simulation and reduce much of the stochastic dynamics of the queueing process to the numerical integration of four differential equations, which is very quick to solve.

1.2. Organization of the Paper

The rest of the paper continues as follows. In Section 2, we review our queueing model and provide expressions for the functional Kolmogorov forward equations for our queueing model. In Section 3, we apply the Gram-Charlier expansion to the functional forward equations and show how this combination improves the estimates of first four cumulant moments of the queue length process. In Section 4, we illustrate that our new techniques are also relevant for constructing accurate estimates of the blocking probability. In Section 5, we provide additional numerical examples to illustrate that our approximations are indeed accurate and good. We also compare the Gram-Charlier method with the method of [10] and show that the Gram-Charlier method is better than the Hermite expansion in [10]. In Section 6, we conclude the paper and give final remarks. Lastly, in the Appendix we provide the proofs of our main theorems and lemmas that are needed in the paper.

2. NONSTATIONARY LOSS QUEUEING MODEL WITH ABANDONMENT

In order to describe the stochastic model for the nonstationary loss queue, we begin with the functional version of the Kolmogorov forward equations for the queue length process. Since our queueing process is an example of a birth-death process with state-dependent rates, we have the following expression for the forward equations of the $M_t/M_t/C_t/K_t/+M_t$ queueing process:

$$\begin{aligned}\dot{E}\left[f(Q)\right] &= \lambda \cdot E\left[\left(f(Q+1)-f(Q)\right)\cdot\{Q<c+k\}\right] \\ &\quad + \mu \cdot E\left[(Q\wedge c)\cdot\left(f(Q-1)-f(Q)\right)\right] \\ &\quad + \beta \cdot E\left[(Q-c)^{+}\cdot\left(f(Q-1)-f(Q)\right)\right], \end{aligned} \tag{2.1}$$

for all integrable functions f. We will always assume, for the remainder of the paper, that quantities such as β and μ are constant. However, the quantities such as β and μ do not have to be constant and our methods work well when the parameters are also functions of

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

time. To simplify our notation, time-dependent quantities such as $Q(t)$, $\lambda(t)$, $c(t)$ and $k(t)$ are denoted in this paper as Q, λ, c, and k, with their time dependence suppressed. For an expression like $E\left[f(Q(t))\right]$ we use the "dot" notation of physics to denote its time derivative when we do not make time explicit or

$$\dot{E}\left[f(Q)\right] \equiv \frac{d}{dt}E\left[f(Q(t))\right]. \tag{2.2}$$

Using special cases of f we can then obtain the following set of Kolmogorov forward equations for the first four cumulant moments

$$\begin{aligned}
\dot{E}[Q] &= \lambda \cdot E[\{Q < c+k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q-c)^+],\\
\dot{\mathrm{Var}}[Q] &= \lambda \cdot E[\{Q < c+k\}] + \mu \cdot E[Q \wedge c] + \beta \cdot E[(Q-c)^+]\\
&\quad + 2\left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] - \mu \cdot \mathrm{Cov}[Q, Q \wedge c] - \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right),\\
\dot{\mathrm{C}}^{(3)}[Q] &= \lambda \cdot E[\{Q < c+k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q-c)^+]\\
&\quad + 3\left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] + \mu \cdot \mathrm{Cov}[Q, Q \wedge c] + \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right)\\
&\quad + 3\left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] - \mu \cdot \mathrm{Cov}\left[\overline{Q}^2, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[\overline{Q}^2, (Q-c)^+\right]\right),\\
\dot{C}^{[4]}[Q] &= \lambda \cdot E[\{Q < c+k\}] + \mu \cdot E\left[Q \wedge c\right] + \beta \cdot E\left[(Q-c)^+\right]\\
&\quad + 4 \cdot \left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] - \mu \cdot \mathrm{Cov}\left[Q, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[Q, (Q-c)^+\right]\right)\\
&\quad + 6 \cdot \left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] + \mu \cdot \mathrm{Cov}\left[\overline{Q}^2, Q \wedge c\right] + \beta \cdot \mathrm{Cov}\left[\overline{Q}^2, (Q-c)^+\right]\right)\\
&\quad + 4 \cdot \left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^3, \{Q < c+k\}\right] - \mu \cdot \mathrm{Cov}\left[\overline{Q}^3, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[\overline{Q}^3, (Q-c)^+\right]\right)\\
&\quad + 12 \cdot \left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] + \mu \cdot \mathrm{Cov}[Q, Q \wedge c] + \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right) \cdot \mathrm{Var}[Q].
\end{aligned}$$

where $\overline{Q} \equiv Q - E[Q]$.

In developing our approximations for the nonstationary loss queue, we will compare our estimates with simulations to judge the accuracy of our approximation methods. To this end, the primary numerical example that we study in this paper to demonstrate the usefulness of our approximation methods has an arrival rate of $\lambda(t) = 10 + 5\sin(t)$, a service rate of $\mu = 1$, an abandonment rate of $\beta = 0.5$, $c = 10$ servers, and $k = 5$ waiting spaces. Moreover, we simulate our queueing model over the time interval (0,40] for 40,000 independent sample paths.

On the left of Figure 1 is a plot of the simulated mean $E[Q(t)]$ and variance $\mathrm{Var}[Q(t)]$ of our queueing process. On the right side of Figure 1, we plot the simulated values of the skewness $\mathrm{Skew}[Q(t)]$ and kurtosis $\mathrm{Kur}[Q(t)]$ of the queueing system. The skewness and kurtosis are related to the third and fourth cumulant moments and are given by the formulas

$$\mathrm{Skew}[Q(t)] = \frac{E[(Q(t) - E[Q(t)])^3]}{\mathrm{Var}[Q(t)]^{3/2}} \quad \text{and} \quad \mathrm{Kur}[Q(t)] = \frac{E[(Q(t) - E[Q(t)])^4]}{\mathrm{Var}[Q(t)]^2} - 3. \tag{2.3}$$

As the right of Figure 1 shows, the skewness and kurtosis are non-zero quantities. Since the skewness and kurtosis for a Gaussian random variable are defined to be zero, on

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

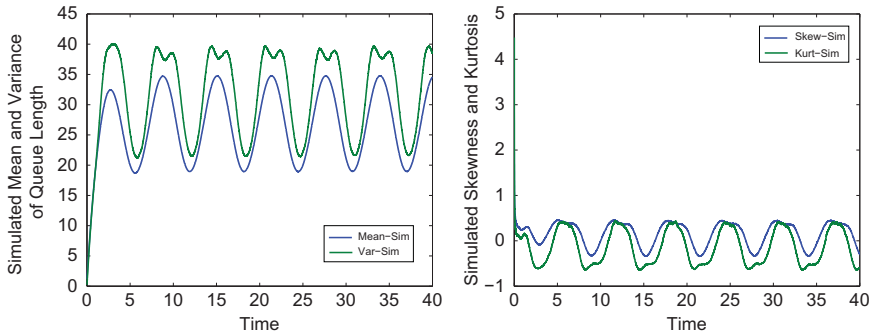

FIGURE 1. (Color online) Simulation of mean and variance of the queueing process (left). Simulation of skewness and kurtosis of the queueing process (right).

the right of Figure 1 gives us supporting evidence that the queueing process distribution is non-Gaussian. However, one also observes from Figure 1 that while the skewness and kurtosis are non-zero, they are also not extremely large quantities either. Since they are not large, this gives us some confidence that using asymptotic expansions around a Gaussian distribution might be reasonable. Moreover, the skewness and kurtosis have the potential to tell us valuable information about the properties of our queueing distribution. In fact, when comparing to a Gaussian distribution, the skewness can tell us whether the median of the queueing distribution is to the left or right of the mean of the distribution and the kurtosis can provide information on the peakedness of the distribution. The skewness is especially important since, the real queueing process is non-negative, unbounded and asymmetric, while the Gaussian distribution can realize negative values and is symmetric around the mean. Thus, the skewness is critical in capturing asymmetries of the queueing distributions. Although the skewness and kurtosis are important statistical and mathematical quantities, they also have some practical value because they can help managers adjust or refine the staffing levels appropriately according the information to the values of the skewness and kurtosis. In fact when the skewness and kurtosis are near zero, they validate the use of the Gaussian approximations. However, when they are away from zero, they can serve to refine Gaussian behavior predicted from rigorous limit theorems.

Unlike the multi-server case with no loss of arrivals, the arrival process of the nonstationary loss queue is *state-dependent*. In fact, the state dependence is not only nonlinear, but it is also discontinuous with respect to the queue length process. This discontinuous nature of the state-dependent arrival rate function precludes the limit theorems of Mandelbaum et al. [8] from being exploited. Thus, it is an important area of research to find new methods for approximating the queue length process, its moment behavior, and various performance measures such as the probability of blocking all at the same time. In the sequel, we present four new methods to use for approximating the dynamics of the nonstationary loss queue.

3. NEW APPROXIMATION METHODS

3.1. Deterministic Mean Approximation

In this section, we give the first approximation for our nonstationary loss queue. It is a purely deterministic method and as a result we define it as the *Deterministic mean approximation* (DMA). The DMA is constructed by assuming $\{q(t)|t \geq 0\}$ is a deterministic process that approximates the queueing process. Thus, we assume that $Q \approx q$ and substitute q for Q in

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

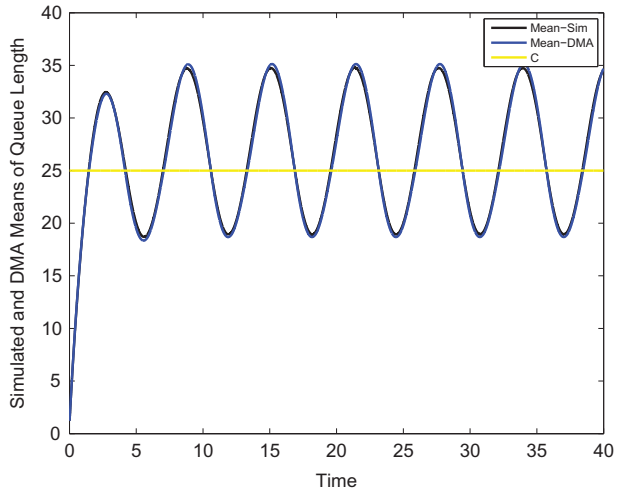

FIGURE 2. (Color online) Simulation of mean and DMA approximation.

the Kolmogorov forward equations for the mean dynamics. As a result, the time derivative of the mean solves the following autonomous differential equation:

$$\dot{q} = \lambda \cdot \{q < c + k\} - \mu \cdot (q \wedge c) - \beta \cdot (q - c)^{+}. \tag{3.1}$$

In Figure 2, we see that the DMA method approximates the mean dynamics of the queue length process fairly well. Since the DMA method is deterministic method, it does not recognize the stochastic fluctuations of the queue length process. Thus, there is a large difference between the DMA and the simulation at the peak of the DMA method. This is because we implicitly assume that all other cumulant moments of the queueing process are negligible and the DMA is unable to use other distributional behavior other than the mean in order to estimate the dynamics of the queue length process. The implicit assumption that all other cumulant moments are negligible is not realistic in practice and warrants a refinement to include more information about the distribution of the queueing process.

3.2. Gaussian Variance Approximation

Our first refinement to the DMA is to assume that our queueing process has a finite variance or second cumulant moment, but all other cumulant moments of order higher than three are assumed to be negligible. Thus, we assume that our queueing model follows a Gaussian distribution. We define this new approximation as the *Gaussian variance approximation* (GVA). This approximation technique was first developed by Massey and Pender [9,10] and Pender [13], which was shown to be equivalent to the method of Ko and Gautam [7]. In [10] and in this paper, we assume that

$$Q(t) \stackrel{d}{=} q(t) + X \cdot \sqrt{v(t)} \tag{3.2}$$

for all $t \geq 0$, where $\{q(t), v(t) | t \geq 0\}$ is some two-dimensional dynamical system where the v process is always positive and X is a standard Gaussian random variable. We also define φ and Φ to be the density and the cumulative distribution functions, for X respectively,

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

where

$$\varphi(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) \equiv \int_{-\infty}^{x} \varphi(y)\, dy, \quad \text{and} \quad \overline{\Phi}(x) \equiv 1 - \Phi(x) = \int_{x}^{\infty} \varphi(y)\ dy. \tag{3.3}$$

THEOREM 3.1: *Suppose that we substitute Eq. 3.2 for the queue length process into the functional forward equations as the surrogate distribution, then the forward equations for the mean and variance of Q are*

$$\dot{E}[Q] = \lambda \cdot E[\{Q < c + k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q - c)^+], \tag{3.4}$$

$$\begin{aligned} \dot{\text{Var}}[Q] = {} & \lambda \cdot E[\{Q < c + k\}] + \mu \cdot E[Q \wedge c] + \beta \cdot E[(Q - c)^+] \\ & + 2\left(\lambda \cdot \text{Cov}[Q, \{Q < c + k\}] - \mu \cdot \text{Cov}[Q, Q \wedge c] - \beta \cdot \text{Cov}[Q, (Q - c)^+]\right), \end{aligned} \tag{3.5}$$

where the unknown expectation and covariance terms have the following values:

$$\begin{aligned} E\left[\{Q < c + k\}\right] &= \Phi(\psi), \\ E\left[(X - \chi)^+\right] &= \phi(\chi) - \chi \cdot \overline{\Phi}(\chi), \\ E\left[(X \wedge \chi)\right] &= \chi \cdot \overline{\Phi}(\chi) - \phi(\chi), \\ \text{Cov}\left[X, \{Q < c + k\}\right] &= \phi(\psi), \\ \text{Cov}\left[X, (X - \chi)^+\right] &= \phi(\chi) - \chi \cdot \overline{\Phi}(\chi), \\ \text{Cov}\left[X, (X \wedge \chi)\right] &= \Phi(\chi), \end{aligned}$$

and where the variable χ and ψ have the values

$$\begin{aligned} \chi &= \frac{c - q}{\sqrt{v}}, \\ \psi &= \frac{c + k - q}{\sqrt{v}}. \end{aligned}$$

Unlike the DMA, the GVA forward equations for the mean also depend on the variance behavior. In this sense the two-dimensional system of the equations for the mean and variance are fully coupled to one another. Thus, now the dynamics of the mean can capture information about the queueing distribution from the variance unlike in the DMA method. Thus, we expect that the dynamics of the mean behavior of the GVA should be different and better than the DMA method.

On the left of Figure 3, we see that the GVA estimate for the mean dynamics is better than the DMA method. This is especially true when the mean queue length peaks. Moreover, on the right of Figure 3, we see that the GVA method is doing a good job of estimating the variance of the queue length process. Looking more closely, we see that the GVA method only does not approximate the dynamics of the queueing process well when skewness and kurtosis reach their local maximums. Thus, one should suspect that the queueing process is the least Gaussian during those times when the skewness and kurtosis are at their local maximums.

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

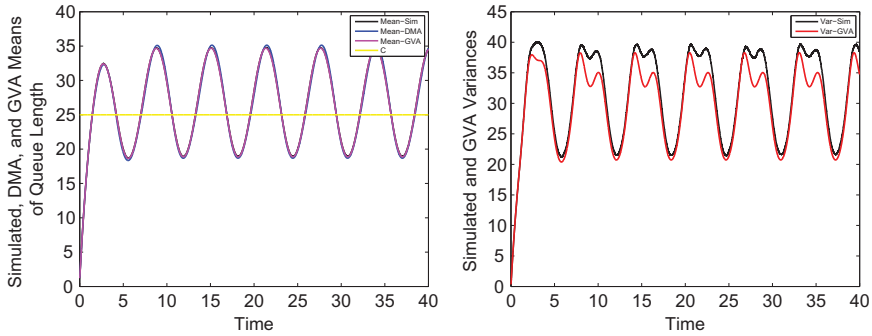

FIGURE 3. (Color online) Simulated, DMA, and GVA means, (left). Simulated and GVA Variances (right).

3.3. Gram-Charlier Skewness Approximation

In this section, we extend the GVA method to include information about the skewness of the queue length process. Following the method developed by Pender [13] for multi-server queues, we assume that the queue length process has the following approximate density:

$$\phi_{\text{Skew}}(x) = \phi(x) \cdot \left(1 + \frac{\kappa_3}{3! \cdot \sqrt{v^3}} \cdot h_3(x)\right) = \phi_{\text{GVA}}(x) + \phi_{\text{GCS}}(x), \tag{3.6}$$

where $\{q, v, \kappa_3\}$ are the mean, variance, and third cumulant moment of the queueing process and $h_3(x)$ is a Hermite polynomial of order 3. Like in the work of [13], we call this approximation the Gram-Charlier Skewness Approximation. We shall show that the skewness allows us to better estimate the mean and variance dynamics of the queueing system with our next theorem:

THEOREM 3.2: *Suppose that Eq. (3.6) is the density for our nonstationary loss queue, then we have the following equations for the mean, variance, and third cumulant moment of nonstationary loss queue with abandonment*

$$\begin{aligned}
\dot{E}[Q] &= \lambda \cdot E[\{Q < c + k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q - c)^+], \\
\dot{\text{Var}}[Q] &= \lambda \cdot E[\{Q < c + k\}] + \mu \cdot E[Q \wedge c] + \beta \cdot E[(Q - c)^+] \\
&\quad + 2\left(\lambda \cdot \text{Cov}[Q, \{Q < c + k\}] - \mu \cdot \text{Cov}[Q, Q \wedge c] - \beta \cdot \text{Cov}[Q, (Q - c)^+]\right), \\
\dot{C}^{(3)}[Q] &= \lambda \cdot E[\{Q < c + k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q - c)^+] \\
&\quad + 3\left(\lambda \cdot \text{Cov}[Q, \{Q < c + k\}] + \mu \cdot \text{Cov}[Q, Q \wedge c] + \beta \cdot \text{Cov}[Q, (Q - c)^+]\right) \\
&\quad + 3\left(\lambda \cdot \text{Cov}\left[\overline{Q}^2, \{Q < c + k\}\right] - \mu \cdot \text{Cov}\left[\overline{Q}^2, Q \wedge c\right] - \beta \cdot \text{Cov}\left[\overline{Q}^2, (Q - c)^+\right]\right),
\end{aligned}$$

where we have the following expressions for the unknown expectations and covariances:

$$\begin{aligned}
E[\{Q < c + k\}] &= \Phi(\psi) - \frac{\kappa_3}{3! \cdot \sqrt{v^3}} \cdot (\psi^2 - 1) \cdot \phi(\psi), \\
E\left[(Q \wedge c)\right] &= q - \sqrt{v} \cdot \phi(\chi) + \chi \cdot \sqrt{v} \cdot \overline{\Phi}(\chi) - \frac{\chi \cdot \phi(\chi) \cdot \kappa_3}{6 \cdot v},
\end{aligned}$$

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

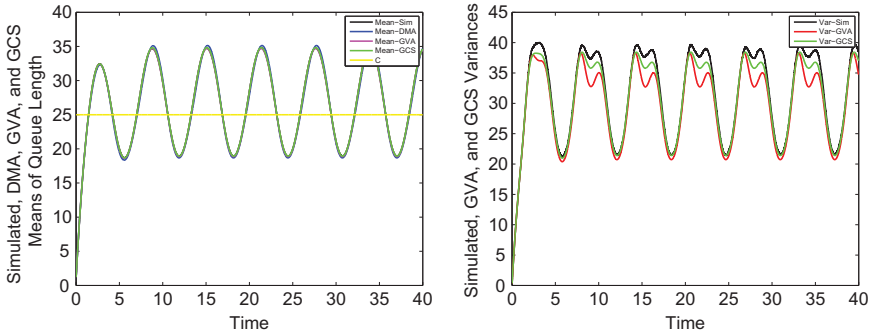

FIGURE 4. (Color online) Simulated, DMA, GVA, GCS means (left). Simulated, GVA, and GCS variances (right).

$$
\begin{aligned}
E\left[(Q-c)^{+}\right] &= \sqrt{v}\cdot\phi(\chi)-\chi\cdot\sqrt{v}\cdot\overline{\Phi}(\chi)+\frac{\chi\cdot\phi(\chi)\cdot\kappa_3}{6\cdot v},\\
\operatorname{Cov}\left[Q,\{Q<c+k\}\right] &= -\sqrt{v}\cdot\phi(\psi)-\frac{\kappa_3}{6\cdot v}\cdot\left(h_3(\psi)+3\cdot h_1(\psi)\right)\cdot\phi(\psi),\\
\operatorname{Cov}\left[Q,(Q-c)^{+}\right] &= v\cdot\overline{\Phi}(\chi)+\frac{(\chi^2+2)\cdot\phi(\chi)\cdot\kappa_3}{6\sqrt{v}},\\
\operatorname{Cov}\left[Q,(Q\wedge c)\right] &= \operatorname{Cov}\left[Q,Q\right]-\operatorname{Cov}\left[Q,(Q-c)^{+}\right],\\
\operatorname{Cov}\left[\overline{Q}^2,\{Q<c+k\}\right] &= -v\cdot h_1(\psi)\cdot\phi(\psi)-\frac{\kappa_3}{6\cdot\sqrt{v}}\cdot\left(h_4(\psi)+7\cdot h_2(\psi)+6\right)\cdot\phi(\psi),\\
\operatorname{Cov}\left[\overline{Q}^2,(Q-c)^{+}\right] &= \sqrt{v^3}\cdot\phi(\chi)+\frac{\kappa_3}{6}\cdot\left[(\chi^3+4\cdot\chi)\cdot\phi(\chi)+6\cdot\overline{\Phi}(\chi)\right],\\
\operatorname{Cov}\left[\overline{Q}^2,(Q\wedge c)\right] &= \kappa_3-\sqrt{v^3}\cdot\phi(\chi)-\frac{\kappa_3}{6}\cdot\left[(\chi^3+4\cdot\chi)\cdot\phi(\chi)+6\cdot\overline{\Phi}(\chi)\right].
\end{aligned}
$$

PROOF: See the Appendix. ■

On the left of Figure 4, we see that the GCS estimate for the mean dynamics is better than the GVA and DMA methods. Once again where the queue length peaks, we have the most improvement of the GCS method over the GVA and DMA methods. Moreover, on the right of Figure 4, we see that the GCS method does a better job of estimating the variance of the queue length process than the GVA method. In fact, the places where the skewness peaks is where there is the most improvement of the variance for the GCS method. This behavior can be confirmed in Figure 5, where we plot the log-relative error of the various approximations. On the left of Figure 5 we have the log-relative error of the mean and it is clear that the GCS method does a better job of estimating the mean dynamics. On the right side of Figure 5 we see that the GCS method is doing a better job of estimating the variance as well. Lastly, in Figure 6, we plot the simulated skewness with the approximation from the GCS method. It is clear that the GCS method doing well at approximating the skewness dynamics. The quality of the skewness approximation is also given on the right of Figure 6, where we plot the log-relative error of the GCS approximation. Overall, it is clear that the GCS method is superior at approximating the time-varying dynamics the nonstationary loss queue. Perhaps, adding more information about the distributional behavior of the queueing process, might add more insight and yield even better approximations.

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

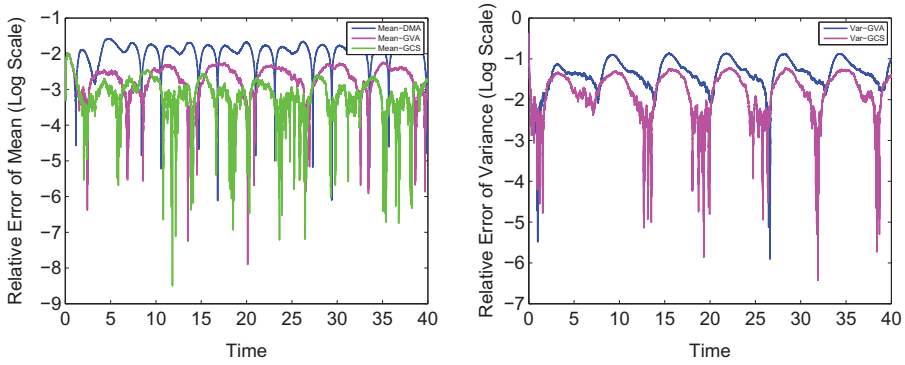

FIGURE 5. (Color online) Log-relative error of DMA, GVA, and GCS means (left). Log-relative error of GVA and GCS variances (right).

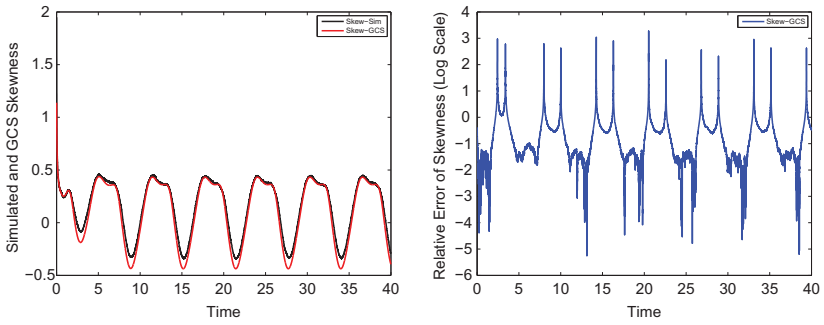

FIGURE 6. (Color online) Simulated and GCS skewness (left). Log-relative error of GCS skewness (right).

3.4. Gram-Charlier Kurtosis Approximation

For this section, we again add another term to the our Gram-Charlier expansion to capture the kurtosis of the nonstationary loss queue. We call this new approximation the Gram-Charlier kurtosis approximation (GCK). Similar to the GCS method, we hope that adding another term will further refine our approximations for the mean, variance, skewness of the queueing model. This will help us attain even better estimates for the mean, variance, and skewness, which can used for better staffing and optimization purposes. For the GCK method, we assume that our queueing process has the following approximate density:

$$\begin{aligned}\phi_{\mathrm{Kur}}(x) &= \phi(x)\cdot\left(1+\frac{\kappa_3}{3!\cdot\sqrt{v^3}}\cdot h_3(x)+\frac{\kappa_4}{4!\cdot v^2}\cdot h_4(x)\right) \\ &= \phi_{\mathrm{GVA}}(x)+\phi_{\mathrm{GCS}}(x)+\phi_{\mathrm{GCK}}(x).\end{aligned} \tag{3.7}$$

Using the GCK approximation as the model for our queueing dynamics allows us to give our next main approximation result.

THEOREM 3.3: *Using the approximate density 3.7, we have the following equations for the mean, variance, third cumulant moment, and fourth cumulant moment of our nonstationary*

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

loss queue with abandonment

$$\dot{E}[Q] = \lambda \cdot E[\{Q < c+k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q-c)^+],$$

$$\begin{aligned}\dot{\mathrm{Var}}[Q] &= \lambda \cdot E[\{Q < c+k\}] + \mu \cdot E[Q \wedge c] + \beta \cdot E[(Q-c)^+] \\ &\quad + 2\left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] - \mu \cdot \mathrm{Cov}[Q, Q \wedge c] - \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right),\end{aligned}$$

$$\begin{aligned}\dot{C}^{(3)}[Q] &= \lambda \cdot E[\{Q < c+k\}] - \mu \cdot E[Q \wedge c] - \beta \cdot E[(Q-c)^+] \\ &\quad + 3\left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] + \mu \cdot \mathrm{Cov}[Q, Q \wedge c] + \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right) \\ &\quad + 3\left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] - \mu \cdot \mathrm{Cov}\left[\overline{Q}^2, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[\overline{Q}^2, (Q-c)^+\right]\right),\end{aligned}$$

$$\begin{aligned}\dot{C}^{[4]}[Q] &= \lambda \cdot E[\{Q < c+k\}] + \mu \cdot E\left[Q \wedge c\right] + \beta \cdot E\left[(Q-c)^+\right] \\ &\quad + 4 \cdot \left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] - \mu \cdot \mathrm{Cov}\left[Q, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[Q, (Q-c)^+\right]\right) \\ &\quad + 6 \cdot \left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] + \mu \cdot \mathrm{Cov}\left[\overline{Q}^2, Q \wedge c\right] + \beta \cdot \mathrm{Cov}\left[\overline{Q}^2, (Q-c)^+\right]\right) \\ &\quad + 4 \cdot \left(\lambda \cdot \mathrm{Cov}\left[\overline{Q}^3, \{Q < c+k\}\right] - \mu \cdot \mathrm{Cov}\left[\overline{Q}^3, Q \wedge c\right] - \beta \cdot \mathrm{Cov}\left[\overline{Q}^3, (Q-c)^+\right]\right) \\ &\quad + 12 \cdot \left(\lambda \cdot \mathrm{Cov}[Q, \{Q < c+k\}] + \mu \cdot \mathrm{Cov}[Q, Q \wedge c] + \beta \cdot \mathrm{Cov}[Q, (Q-c)^+]\right) \cdot \mathrm{Var}[Q],\end{aligned}$$

where we have the following expressions for the unknown expectations and covariances

$$E[\{Q < c+k\}] = \Phi(\psi) - \frac{\kappa_3}{6 \cdot \sqrt{v^3}} \cdot h_2(\psi) \cdot \phi(\psi) - \frac{\kappa_4}{24 \cdot v^2} \cdot h_3(\psi) \cdot \phi(\psi),$$

$$\begin{aligned}E\left[(Q \wedge c)\right] &= q - \sqrt{v} \cdot \phi(\chi) + \chi \cdot \sqrt{v} \cdot \overline{\Phi}(\chi) - \frac{\chi \cdot \phi(\chi) \cdot \kappa_3}{6 \cdot v} \\ &\quad - \frac{(\chi^2 - 1) \cdot \phi(\chi) \cdot \kappa_4}{6 \cdot \sqrt{v^3}},\end{aligned}$$

$$E\left[(Q-c)^+\right] = \sqrt{v} \cdot \phi(\chi) - \chi \cdot \sqrt{v} \cdot \overline{\Phi}(\chi) + \frac{\chi \cdot \phi(\chi) \cdot \kappa_3}{6 \cdot v} + \frac{(\chi^2 - 1) \cdot \phi(\chi) \cdot \kappa_4}{6 \cdot \sqrt{v^3}},$$

$$\begin{aligned}\mathrm{Cov}\left[Q, \{Q < c+k\}\right] &= -\sqrt{v} \cdot \phi(\psi) - \frac{\kappa_3}{6 \cdot v} \cdot (h_3(\psi) + 3 \cdot h_1(\psi)) \cdot \phi(\psi) \\ &\quad - \frac{\kappa_4}{24 \cdot \sqrt{v^3}} \cdot (h_4(\psi) + 4 \cdot h_2(\psi)) \cdot \phi(\psi),\end{aligned}$$

$$\mathrm{Cov}\left[Q, (Q-c)^+\right] = v \cdot \overline{\Phi}(\chi) + \frac{(\chi^2 + 2) \cdot \phi(\chi) \cdot \kappa_3}{6\sqrt{v}},$$

$$\mathrm{Cov}\left[Q, (Q \wedge c)\right] = v - \mathrm{Cov}\left[Q, (Q-c)^+\right],$$

$$\begin{aligned}\mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] &= v \cdot h_1(\psi) \cdot \phi(\psi) - \frac{\kappa_3}{6 \cdot \sqrt{v}} \cdot (h_4(\psi) + 7 \cdot h_2(\psi) + 6) \cdot \phi(\psi) \\ &\quad - \frac{\kappa_4}{24 \cdot v} \cdot (h_5(\psi) + 9 \cdot h_3(\psi) + 12 \cdot h_1(\psi)) \cdot \phi(\psi),\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\left[\overline{Q}^2, (Q-c)^+\right] &= \sqrt{v^3} \cdot \phi(\chi) + \frac{\kappa_3}{6} \cdot \left[(\chi^3 + 4 \cdot \chi) \cdot \phi(\chi) + 6 \cdot \overline{\Phi}(\chi)\right] \\ &\quad + \frac{\kappa_4}{24 \cdot \sqrt{v}} \cdot (\chi^4 + 3 \cdot \chi^2 + 6) \cdot \phi(\chi),\end{aligned}$$

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

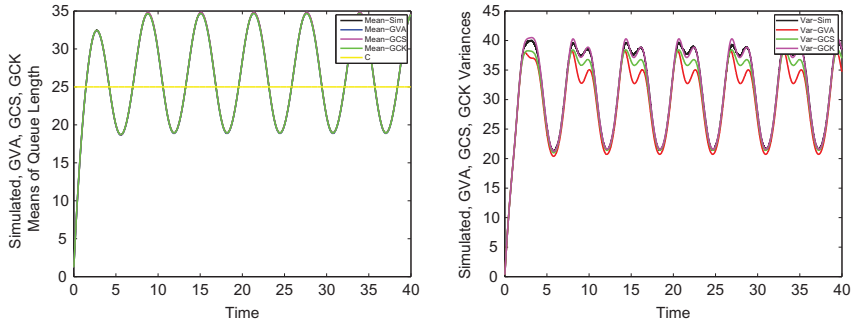

FIGURE 7. (Color online) Simulated, GVA, GCS, and GCK means (left). Simulated, GVA, GCS, and GCK variances (right).

$$\begin{aligned}
\mathrm{Cov}\left[\overline{Q}^2,(Q\wedge c)\right] &= \kappa_3 - \mathrm{Cov}\left[\overline{Q}^2,(Q-c)^+\right],\\
\mathrm{Cov}\left[\overline{Q}^3,\{Q<c+k\}\right] &= \sqrt{v^3}\cdot h_1(\psi)\cdot\phi(\psi) - \frac{\kappa_3}{6}\cdot\left(h_5(\psi)+12\cdot h_3(\psi)+27\cdot h_1(\psi)\right)\cdot\phi(\psi)\\
&\quad - \frac{\kappa_2}{24\cdot\sqrt{v}}\cdot\left(h_6(\psi)+15\cdot h_4(\psi)+48\cdot h_2(\psi)+24\right)\cdot\phi(\psi),\\
\mathrm{Cov}\left[\overline{Q}^3,(Q-c)^+\right] &= v^2\cdot\left(\left(\chi^2+1\right)\cdot\phi(\chi)\right)+3\cdot v^2\cdot\overline{\Phi}(\chi)\\
&\quad + \frac{\kappa_3\cdot\sqrt{v}}{6}\cdot\left(\left(h_4(\chi)+12\cdot h_2(\chi)+27\right)\cdot\phi(\chi)\right)+\frac{\kappa_4}{24\cdot v^2}\\
&\quad \cdot\sqrt{v^3}\cdot\left(\left(h_5(\chi)+15\cdot h_3(\chi)+48\cdot h_1(\chi)\right)\cdot\phi(\chi)+24\cdot\overline{\Phi}(\chi)\right),\\
\mathrm{Cov}\left[\overline{Q}^3,(Q\wedge c)\right] &= 3\cdot v^2+\kappa_4-\mathrm{Cov}\left[\overline{Q}^3,(Q-c)^+\right].
\end{aligned}$$

PROOF: See the Appendix. ■

In Figures 7 and 9, we see that we can estimate the mean, variance, skewness, and kurtosis quite well by adding an additional term to approximate the kurtosis of the queueing distribution. We clearly see that the GCK approximation is doing the best at approximating the mean, variance, and skewness of the queueing process.

4. ESTIMATING BLOCKING PROBABILITIES

The blocking probability is perhaps the most important performance measure of the nonstationary loss queue. Unlike its stationary counterpart, the nonstationary loss queue is not insensitive to the service distribution; see for example Davis et al. [1]. In this section, we give approximations for the blocking probabilities for the nonstationary loss queue.

4.1. GVA Blocking Probability

Using the GVA approximation for the nonstationary loss queueing process we can also derive an approximate formula for the probability of blocking or the probability that customer

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

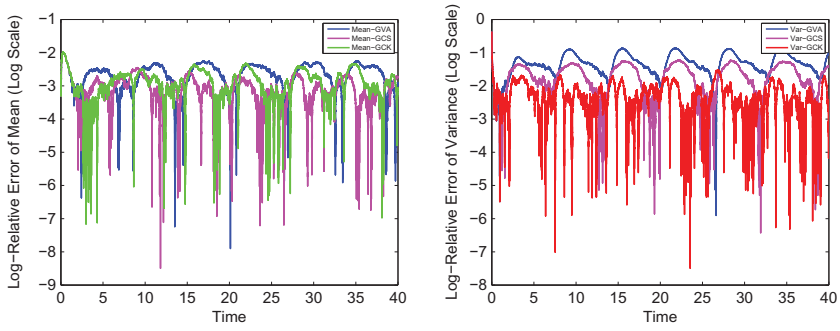

FIGURE 8. (Color online) Log-relative error of GVA, GCS, and GCK means (left). Log-relative error of GVA, GCS, and GCK variances (right).

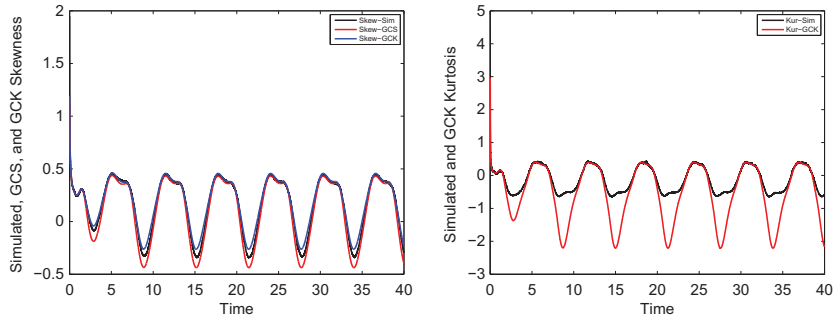

FIGURE 9. (Color online) Simulated, GCS, and GCK skewness (left). Simulated and GCK kurtosis (right).

who enters the queue at time t will be turned away for service. For GVA, the probability of blocking is:

$$\begin{aligned}\mathbb{P}\{Q \geq c+k\} &= \mathbb{P}\{q + X \cdot \sqrt{v} \geq c+k\} \\ &= \mathbb{P}\{X \geq \psi\} \\ &= \overline{\Phi}(\psi).\end{aligned}$$

This formula asserts that we can approximate the probability of blocking with the Gaussian tail distribution and yields some insight on the dynamics of our queueing system's probabilistic behavior.

4.2. GCS Blocking Probability

The GCS approximation like GVA also allows us to calculate the probability of delay. Under the assumptions of the GCS density, the approximate probability of delay is:

$$\begin{aligned}\mathbb{P}\{Q \geq c+k\} &= E_{\text{Skew}}[\{X \geq \psi\}] \\ &= E_{\text{GVA}}[\{X \geq \psi\}] + E_{\text{GCS}}[\{X \geq \psi\}] \\ &= \overline{\Phi}(\psi) + (\psi^2 - 1) \cdot \phi(\psi) \cdot \frac{\kappa_3}{6 \cdot \sqrt{v^3}}.\end{aligned}$$

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

Like the GCS density 3.6, the GCS approximation for the blocking probability is a perturbation of the probability of blocking of the GVA, which includes the skewness. Thus, if the skewness is zero, we obtain the GVA blocking probability as a special case. Moreover, if one looks more closely at the GCS approximation for the blocking probability, one notices that the skewness correction term provided from the Gram-Charlier expansion is exactly the second-order Edgeworth expansion term.

4.3. GCK Blocking Probability

The GCK approximation also allows us to calculate many probabilistic quantities of interest like the the probability of delay. For GCK, the probability of delay is:

$$\begin{aligned}\mathbb{P}\{Q \geq c+k\} &= E_{\mathrm{Kur}}[\{X \geq \psi\}] \\ &= E_{\mathrm{GVA}}[\{X \geq \psi\}] + E_{\mathrm{GCS}}[\{X \geq \psi]\}] + E_{\mathrm{GCK}}[\{X \geq \psi\}] \\ &= \overline{\Phi}(\chi) + (\psi^2 - 1) \cdot \phi(\psi) \cdot \frac{\kappa_3}{6 \cdot \sqrt{v^3}} + (\psi^3 - 3 \cdot \psi) \cdot \phi(\psi) \cdot \frac{\kappa_4}{24 \cdot v^2}.\end{aligned}$$

Like the density, the GCK approximation for the blocking probability is a perturbation of the blocking probability of the GCS, which includes the kurtosis. Thus, if the kurtosis is zero, we get back the GCS probability of blocking. Similar to the GCS approximation for the blocking probability, the kurtosis correction term provided from the GCK expansion is exactly the third order Edgeworth expansion term.

On the left of Figure 10, we also simulated the blocking probability for the nonstationary loss queue and compared it with the GVA, GCS, and GCK approximations. We see that the GCS method does slightly better than the GVA and GCK approximations. However, the improvement is very slight. This is confirmed on the right of Figure 10 where we plot the log-relative error of the various approximations. Once again we see that the GCS method is superior, but only slightly superior to the GCK method.

4.4. Comparison Against GSA

In this section, we compare the Gram-Charlier expansion with the Gaussian skewness approximation (GSA) of [10] using the parameters of Table 1. As one can see in Figure 11, we see that the GCS approximation is much better than the GSA approximation. This is especially seen in the variance and skewness of the queueing process. Moreover, we see that the GCS is better at estimating the blocking probability as well.

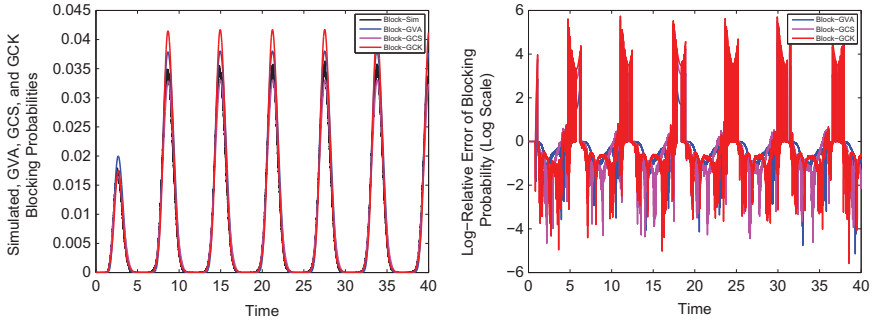

FIGURE 10. (Color online) Simulated, GVA, GCS, and GCK blocking probability (left). Log-relative error of GVA, GCS, and GCK blocking probability (right).

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

TABLE 1. High Arrival Rate Parameters

Parameter	Value
$\lambda(t)$	$100 + 40 \cdot \sin t$
μ	1
$c(t)$	100
$k(t)$	80
β	2

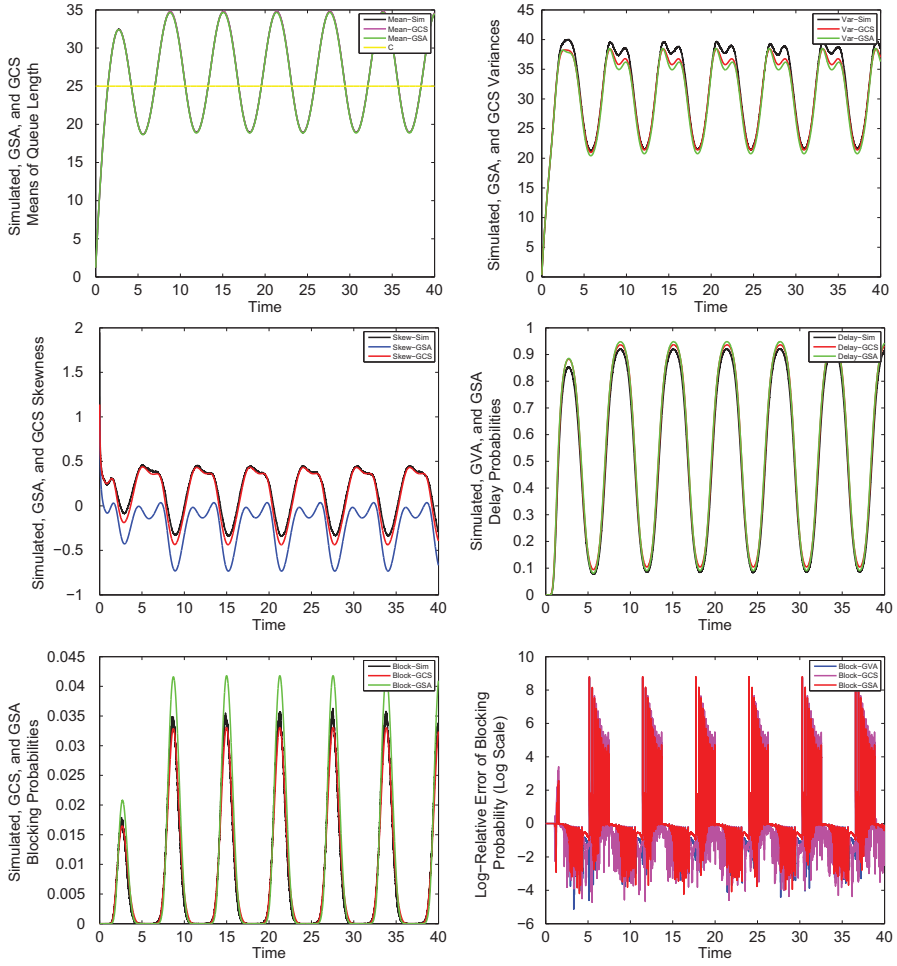

FIGURE 11. (Color online) Comparing the Gram-Charlier and Gaussian skewness methods.

5. ADDITIONAL NUMERICAL EXAMPLES

In this section, we give additional numerical examples of our methods with skewness and kurtosis corrections. These additional examples give support that our new methods work in a variety of parameter settings that are important in practice. Software to implement some of these methods is available on the author's website.

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

5.1. Dynamic Staffing Example

In Figure 12 we give an example of a queueing system, where the number of servers changes dynamically through time. The parameters we use to simulate the loss queue are given in Table 2. On the top left of Figure 12, we see that all of the methods approximate the mean

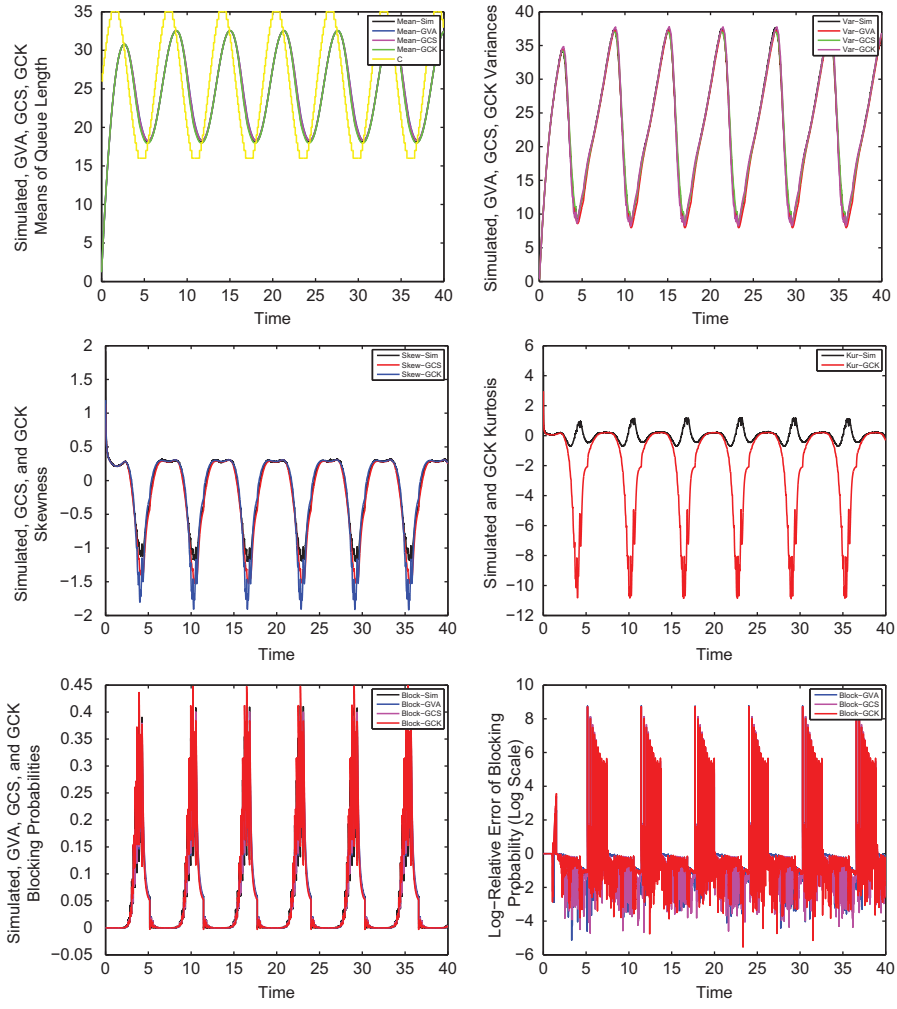

Figure 12. (Color online) Sinusoidal arrival rate and staffing schedule (dynamic staffing example).

Table 2. Dynamic Staffing Parameters

Parameter	Value
$\lambda(t)$	$10 + 5 \cdot \sin t$
μ	1
$c(t)$	$\lceil \lambda(t) \cdot \frac{3}{2} \rceil$
β	0.25

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

TABLE 3. High Arrival Rate Parameters

Parameter	Value
$\lambda(t)$	$100 + 40 \cdot \sin t$
μ	1
$c(t)$	100
$k(t)$	80
β	2

TABLE 4. High Arrival Rate Parameters

Parameter	Value
$\lambda(t)$	$100 + 40 \cdot \sin t$
μ	1
$c(t)$	100
$k(t)$	80
β	0.5

dynamics well. Even for the variance on the top right of Figure 12 it seems that all of the approximations doing quite well at approximating the nonstationary behavior. On the middle left of Figure 12, we see that the GCS method is outperforming the GCK method for estimating the skewness. However, this can be explained because in this example, the kurtosis is not well approximated by the GCK method on the middle right of Figure 12. Moreover, all the methods seem to work well at approximating the blocking probability, which is an important performance measure. However, the GCK does slightly worse than the GCS since the GCK does not accurately approximate the kurtosis well. Thus, this example provides evidence that our Gram-Charlier expansion method is accurate even when the number of servers dynamically changes throughout time and is not just a fixed constant.

5.2. Large-Scale and Impatient Customers

In Figure 13, we give an example of the dynamics of a queueing system with a high arrival rate, a large number of servers, and where the customers are very impatient. The parameters that we use for this example are given in Table 3. We see on the top of Figure 13 that the mean and variance are approximated very well regardless of the method used. One reason is that we are very close to operating in the many server heavy traffic regime and distribution is becoming more *Gaussian like*. On the middle left of Figure 13, we see that the GCS and GCK methods are doing very well at approximating the skewness of the queueing process. Furthermore on the right middle of Figure 13, we see that the GCK method is approximating the kurtosis very accurately. On the bottom left of Figure 13, we see that GVA, GCS, and GCK are all doing a good job of estimating the probability of blocking when there is not much blocking since customers are impatient and abandon the queue quickly if they are forced to wait. With a high arrival rate and large number of servers, it is not complete necessary to use the skewness and kurtosis corrections as there is not much room for correcting the estimates of the mean and variance dynamics since the queueing process mimics an infinite server queue. However, one other thing to notice is that the skewness has a local maximum when the queueing process is critically loaded that is, $(q = c)$ where we expect the queueing system not to behave like a Gaussian process.

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

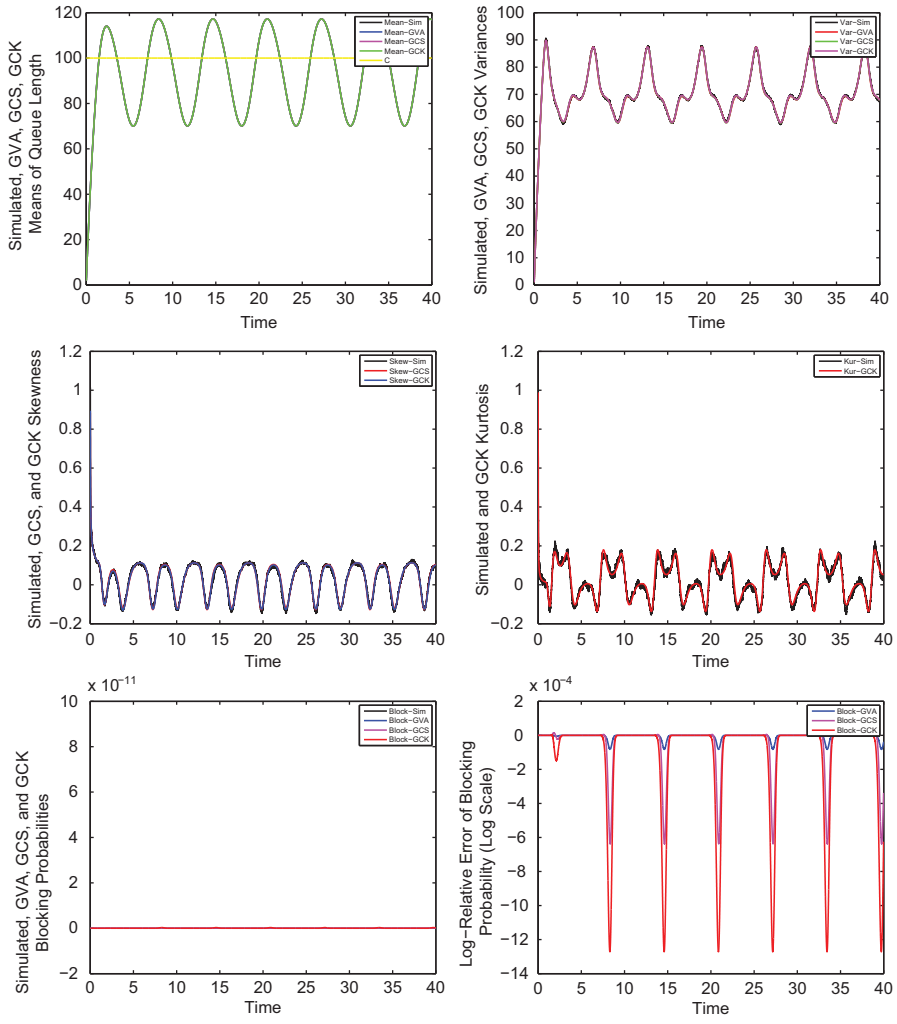

FIGURE **13.** (Color online) Impatient relative to mean service rate example.

5.3. Large-Scale and Patient Customers

In Figure 14, we give an example of the dynamics of a queueing system with a high arrival rate, a large number of servers, and where the customer are relatively patient and are willing to wait for service. The parameters that we use for this example are given in Table 4. On the top of Figure 14 we see that the mean and variance are approximated very well regardless of the method used. However, we see that the GCS and GCK methods are improvements over the GVA method, especially for the variance. On the middle left of Figure 14 we see that the GCS and GCK methods are doing very well at approximating the skewness of the queueing process. Once again we see that the GCK method is better at approximating the skewness since it incorporates more information about the queueing process. Furthermore on the right middle of Figure 14, we see that the GCK method is approximating the kurtosis very accurately. On the bottom of Figure 14, we see that the GCK method is the best at approximating the probability of blocking. In fact, we see consistent improvment of the Gram-Charlier method as we add more terms in the expansion.

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

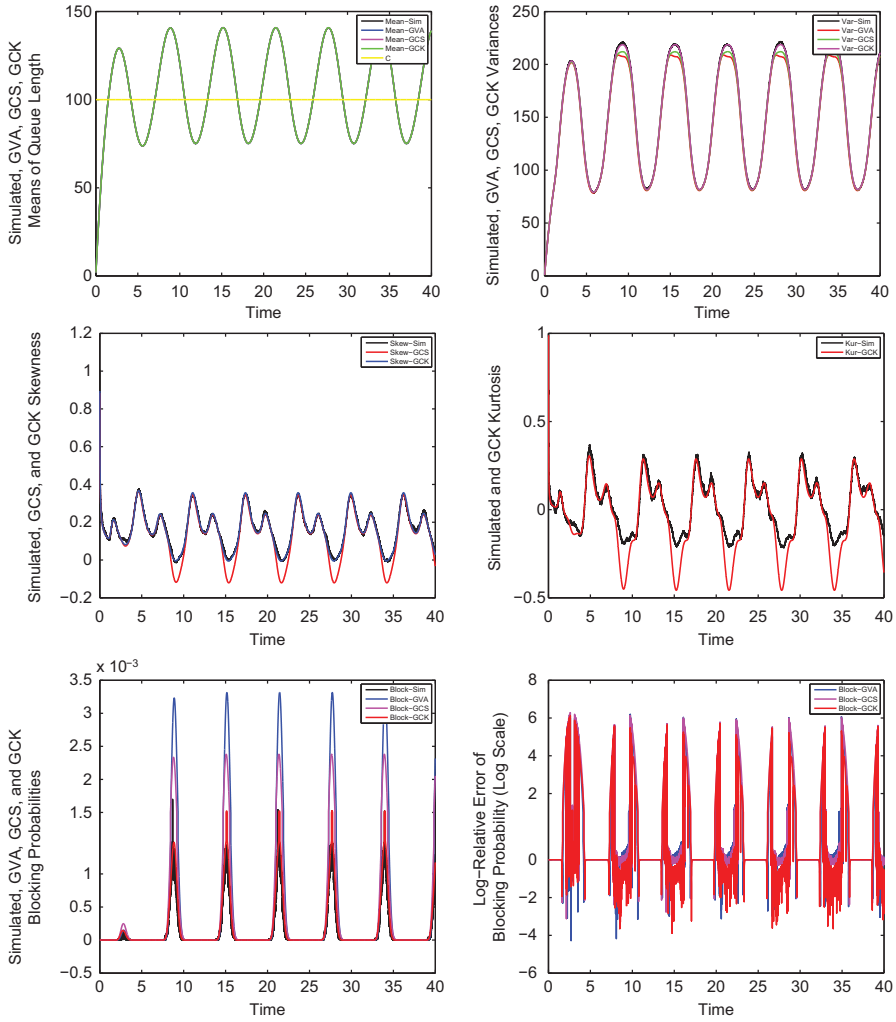

FIGURE 14. (Color online) Patient relative to mean service rate example.

5.4. Additional Comparison Against GSA

In this section, we provide an additional example to compare the Gram-Charlier expansion with the GSA of [10]. Using the parameters of Table 5 one can see in Figure 11 that the GCS approximation is better than the GSA approximation. This is especially seen in the

TABLE 5. High Arrival Rate Parameters

Parameter	Value
$\lambda(t)$	$100 + 40 \cdot \sin t$
μ	1
$c(t)$	100
$k(t)$	80
β	0.5

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

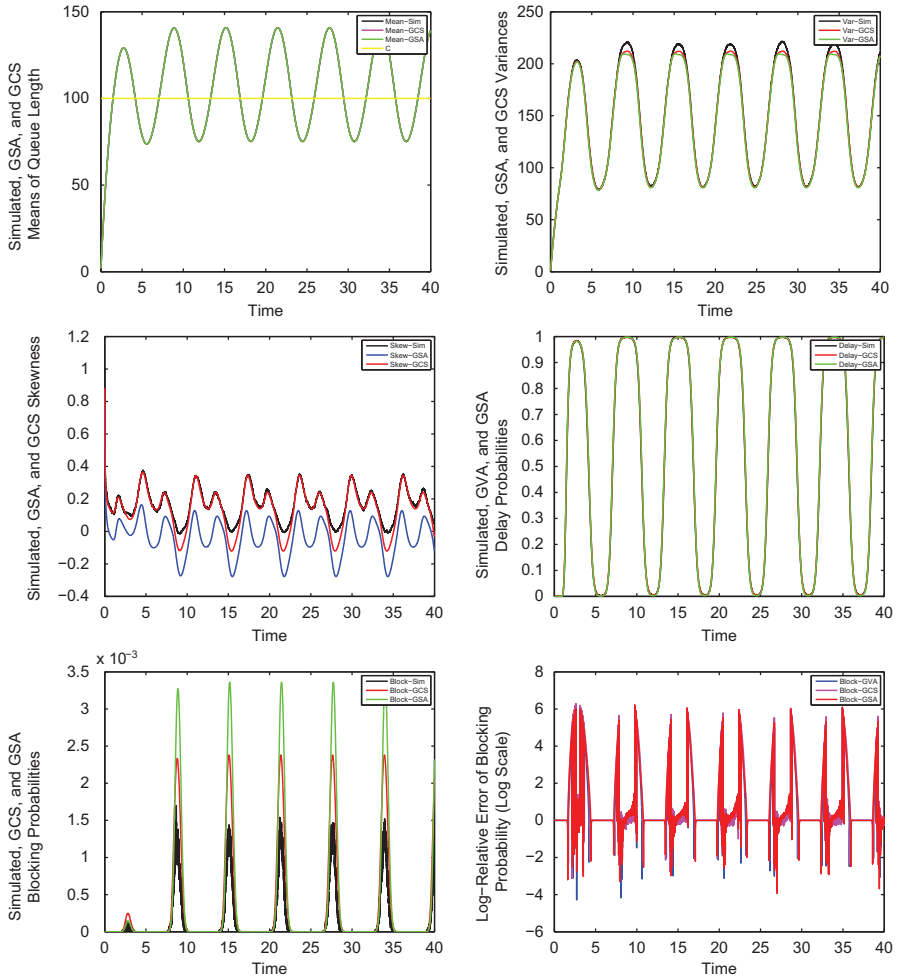

FIGURE 15. (Color online) Comparing the Gram-Charlier and Gaussian skewness methods (second example).

variance and skewness of the queueing process. Moreover, we see that the GCS is better at estimating the blocking probability as well. Thus, this shows that we should use the Gram-Charlier method in the nonstationary loss case since it is more accurate and the analysis is much simpler since it does not use polynomial roots. Moreover, we are also able to capture higher moments with less effort using the Gram-Charlier method.

6. CONCLUSION

In this paper, we have illustrated that combining the functional Kolmogorov forward equations with the Gram-Charlier series expansion is a good method for approximating the dynamics of the nonstationary loss queue with abandonment. Thus, this method can be applied to other queueing processes that are not a part of the Markovian service network family. The Gram-Charlier approach generates a finite-dimensional dynamical system that improves our estimation of both the mean and variance of the original queueing process

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

and most of the time also estimates the skewness and kurtosis fairly well. This is especially needed during the times of critical loading or when the loss queue experiences significant blocking. Estimation of the blocking probability is perhaps the most important performance measure of the nonstationary loss queue and our method demonstrates that it can accurately reproduce the simulated results. We hope to extend this method to a network of loss queues, which would be useful for modeling networks of service systems that mimic the behavior of the loss queue.

More recently, Pender [14] has used Laguerre polynomials for expanding the queue length process of multiserver queues with small numbers of servers. A comparison of the two methods for nonstationary loss queues is a subject of future research.

References

1. Davis, J.L., Massey, W.A. & Whitt, W. (1995). Sensitivity to the service-time distribution in the nonstationary Erlang loss model. *Management Science* 41: 1107–1116.
2. Grier, N., Massey, W.A., McKoy, T. & Whitt, W. (1997). The time-dependent Erlang loss model with retrials. *Telecommunication Systems* 7: 253–265.
3. Hampshire, R. (2007). *Dynamic Queueing Models for the Operations Management of Communication Services*, Ph.D. Thesis, Princeton University.
4. Hampshire, R. Jennings, O.B. & Massey, W.A. (2009). A time varying call center design with Lagrangian mechanics. *Probability in the Engineering and Informational Sciences* 23: 231–259.
5. Hampshire, R. & Massey, W.A. (2010). A tutorial on dynamic optimization and applications to queueing systems with time-varying rates. *Tutorials in Operations Research* 208–247.
6. Jagerman, D.L. (1975). Nonstationary blocking in telephone traffic. *Bell System Technical Journal* 54: 625–661.
7. Ko, Y.M. & Gautam, N. (2013). Critically loaded time-varying multiserver queues: computational challenges and approximations. *Informs Journal on Computing* 25: 285–301.
8. Mandelbaum, A., Massey, W.A. & Reiman, M. (1998). Strong approximations for Markovian service networks. *Queueing Systems*, 30: 149–201.
9. Massey, W.A. & Pender, J. (2011). Skewness variance approximation for dynamic rate multi-server queues with abandonment. *ACM SIGMETRICS Performance Evaluation Review* 39: 74–74.
10. Massey, W.A. & Pender, J. (2013). Gaussian skewness approximation for dynamic rate multiserver queues with abandonment. *Queueing Systems* 75: 243–277.
11. Massey, W.A. & Whitt, W. (1994). An analysis of the modified offered load approximation for the Erlang loss model. *Annals of Applied Probability* 4: 1145–1160.
12. Massey, W.A. & Whitt, W. (1996). Stationary-process approximations for the nonstationary Erlang loss model. *Operations Research*, 44: 976–983.
13. Pender, J. (2014). Gram Charlier expansions for time varying multiserver queues with abandonment. *SIAM Journal of Applied Mathematics To Appear*
14. Pender, J. (2012). Time Varying Queues with Abandonment Via Laguerre Polynomial Expansions, Princeton University, http://www.princeton.edu/∼jpender/publications
15. Stein, C.M. (1986). *Approximate computation of expectations*, Lecture Notes Monograph Series, Vol. 7, Hayward, CA: Institute of Mathematical Statistics.
16. Wallace, R. (2004). Performance modeling and design of call centers with skill-based routing. Ph.D. dissertation, School of Engineering and Applied Science, George Washington University, Washington, DC.

APPENDIX A

PROPOSITION A.1: *Any L^2 function can be written as an infinite sum of Hermite polynomials of X, that is,*

$$f(X) \stackrel{L^2}{=} \sum_{n=0}^{\infty} \frac{1}{n!} E[f^{(n)}(X)] \cdot h_n(X)$$

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

and the expectation of two functions of Hermite polynomials has the following decomposition

$$E[f(X)\cdot g(X)]=\sum_{n=0}^{\infty}\frac{1}{n!}\cdot E[f^{(n)}(X)]\cdot E[g^{(n)}(X)].$$

The next lemma known as Stein's lemma [15] is how we calculate many of the expectation and covariance terms with explicit Hermite expressions.

LEMMA A.2 (Stein [15]): *The random variable X is Gaussian $(0,1)$ if and only if*

$$E\left[X\cdot f(X)\right]=E\left[\frac{d}{dX}f(X)\right], \tag{A.1}$$

for all generalized functions f. Moreover,

$$E\left[h_n(X)\cdot f(X)\right]=E\left[\frac{d^n}{dX^n}f(X)\right], \tag{A.2}$$

where $h_n(X)$ is the n^{th} Hermite polynomial.

A.1. Calculations of Unknown Expectations and Covariance Terms

In this section, we derive explicit formulas for the expectations and covariances needed for construct our dynamical system approximation for our queueing process.

A.1.1. Computation of Arrival Function Terms Now we compute the arrival rate function terms using the Stein's lemma and Hermite polynomials representations and properties. We only compute the terms for the GCK method since GCS can be obtained by setting κ_4 to zero. Moreover, GVA terms can be obtained by setting both κ_3 and κ_4 to zero. Thus, it suffices to only compute the expectation and covariance terms only for the case of the GCK.

$$\begin{aligned}
E\left[\{Q<c+k\}\right]&=1-E\left[\{X\geq\psi\}\right]-\frac{\kappa_3}{6\cdot v}\cdot E\left[h_3(X)\cdot\{X\geq\psi\}\right]\\
&\quad-\frac{\kappa_4}{24\cdot\sqrt{v^3}}\cdot E\left[h_4(X)\cdot\{X\geq\psi\}\right]\\
&=\Phi(\psi)-\frac{\kappa_3}{6\cdot\sqrt{v^3}}\cdot h_2(\psi)\cdot\phi(\psi)-\frac{\kappa_4}{24\cdot v^2}\cdot h_3(\psi)\cdot\phi(\psi),\\
\mathrm{Cov}\left[Q,\{Q<c+k\}\right]&=-\mathrm{Cov}\left[Q,\{Q\geq c+k\}\right]\\
&=-\sqrt{v}\cdot\mathrm{Cov}\left[X,\{X\geq\psi\}\right]-\frac{\kappa_3}{6\cdot v}\cdot\mathrm{Cov}\left[X\cdot h_3(X),\{X\geq\psi\}\right]\\
&\quad-\frac{\kappa_4}{24\cdot\sqrt{v^3}}\cdot\mathrm{Cov}\left[X\cdot h_4(X),\{X\geq\psi\}\right]\\
&=-\sqrt{v}\cdot\phi(\psi)-\frac{\kappa_3}{6\cdot v}\cdot(h_3(\psi)+3\cdot h_1(\psi))\cdot\phi(\psi)\\
&\quad-\frac{\kappa_4}{24\cdot\sqrt{v^3}}\cdot(h_4(\psi)+4\cdot h_2(\psi))\cdot\phi(\psi),
\end{aligned}$$

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press

$$\begin{aligned}
\mathrm{Cov}\left[\overline{Q}^2, \{Q < c+k\}\right] &= -\mathrm{Cov}\left[\overline{Q}^2, \{Q \geq c+k\}\right] \\
&= -v \cdot \mathrm{Cov}\left[X^2, \{X \geq \psi\}\right] - \frac{\kappa_3}{6 \cdot \sqrt{v}} \cdot \mathrm{Cov}\left[X^2 \cdot h_3(X), \{X \geq \psi\}\right] \\
&\quad - \frac{\kappa_4}{24 \cdot v} \cdot \mathrm{Cov}\left[X^2 \cdot h_4(X), \{X \geq \psi\}\right] \\
&= -v \cdot h_1(\psi) \cdot \phi(\psi) - \frac{\kappa_3}{6 \cdot \sqrt{v}} \cdot (h_4(\psi) + 7 \cdot h_2(\psi) + 6) \cdot \phi(\psi) \\
&\quad - \frac{\kappa_4}{24 \cdot v} \cdot (h_5(\psi) + 9 \cdot h_3(\psi) + 12 \cdot h_1(\psi)) \cdot \phi(\psi) \\
\mathrm{Cov}\left[\overline{Q}^3, \{Q < c+k\}\right] &= -\mathrm{Cov}\left[\overline{Q}^3, \{Q \geq c+k\}\right] \\
&= -\sqrt{v^3} \cdot \mathrm{Cov}\left[X^3, \{X \geq \psi\}\right] - \frac{\kappa_3}{6} \cdot \mathrm{Cov}\left[X^3 \cdot h_3(X), \{X \geq \psi\}\right] \\
&\quad - \frac{\kappa_4}{24 \cdot \sqrt{v}} \cdot \mathrm{Cov}\left[X^3 \cdot h_4(X), \{X \geq \psi\}\right] \\
&= -\sqrt{v^3} \cdot (\psi^2 + 2) \cdot \phi(\psi) - \frac{\kappa_3}{6} \cdot (h_5(\psi) + 12 \cdot h_3(\psi) + 27 \cdot h_1(\psi)) \cdot \phi(\psi) \\
&\quad - \frac{\kappa_2}{24 \cdot \sqrt{v}} \cdot (h_6(\psi) + 15 \cdot h_4(\psi) + 48 \cdot h_2(\psi) + 24) \cdot \phi(\psi).
\end{aligned}$$

For all the other expectation and covariance terms that are used in calculating the relevant cumulant moments and performance measures, the derivation of those expressions can be found in [13].

https://doi.org/10.1017/S0269964814000205 Published online by Cambridge University Press