Probability in the Engineering and Informational Sciences, **23**, 2009, 261–280. Printed in the U.S.A.
doi:10.1017/S0269964809000163

ON THE TIME-DEPENDENT OCCUPANCY DISTRIBUTION OF THE $G/G/1$ QUEUING SYSTEM

JORGE LIMÓN–ROBLES
Industrial and Systems Engineering
Instituto Tecnológico y de Estudios Superiores de Monterrey
C.P. 64849, Monterrey, N.L., México
E-mail: jorge.limon@itesm.mx

MARTIN A. WORTMAN
Department of Industrial and Systems Engineering
Texas A&M University
College Station, TX 77843–3131
E-mail: wortman@tamu.edu

This article offers an approach for studying the time-dependent occupancy distribution for a modest generalization of the $GI/G/1$ queuing system in which interarrival times and service times, although mutually independent, are *not* necessarily identically distributed. We develop and explore an analytical model leading to a computational approach that gives tight bounds on the occupancy distribution. Although there is no general closed-form characterization of probability law dynamics for occupancy in the $GI/G/1$ queue, our results offer what might be termed "near-closed-form" in that accurate plots of the transient occupancy distribution can be constructed with an insignificant computational burden. We believe that our results are unique; we are unaware of any alternative analytical approach leading to a numerical characterization of the *time-dependent* occupancy distribution for the $G/G/1$ queuing systems considered here.

Our analyses employ a marked point process that converges to the occupancy process at any fixed time t; it is shown that this process forms a Markov chain from which the transient occupancy law is available. We verify our analytical approach via comparison with the well-known closed-form expressions for time-dependent occupancy distribution of the $M/M/1$ queue. Additionally, we suggest the viability of our approach, as a computational means of obtaining the time-dependent occupancy distribution, through straightforward application to a $Gamma^{[x]}/Weibull/1$ queuing system having batch arrivals and batch job services.

© 2009 Cambridge University Press 0269-9648/09 $25.00
https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

1. INTRODUCTION

The $G/G/1$ queuing system remains an important delay-system paradigm having wide application. This paradigm has been much studied since the 1950s, with major emphasis given to the analysis of stationary probability laws on waiting time and occupancy. There exist many important results, including Lindley [7, Sect. 6.1], Kingman [9], Borovkov [2], and Cohen [6], among others. Most of the available analytical results are expressed in forms not necessarily amenable to numerical computation or they are approximations that apply only in stylized special cases.

Historically, approximations of *stationary* moments and probability law have dominated the literature devoted to computational queuing analysis. Seminal works [1,4,5,8,10,13,15,16] have been widely cited and extended. Yet, the body of literature citing these references is devoted to the study of stationary behavior. Results addressing the dynamics of occupancy probability law for the $G/G/1$ paradigm are limited. However, in many practical situations, the knowledge of the stationary probability law alone is insufficient and the dynamics of the occupancy law cannot be ignored. The objective of this article is to offer a numerical approach for quickly and accurately exploring the dynamics of the occupancy distribution for a modest generalization of the $GI/G/1$ queue.

The available results most closely related to our approach appear to be given by Litko [12], who obtained an approximation to the stationary occupancy and interdeparture time distributions for $GI/G/1$ queue. He noted that his approach can be used to obtain the queue length distribution at each arrival epoch in an iterative fashion.

Litko's approach employs Laguerre transforms to map continuous distributions into a discrete sequence of real numbers (the coefficients of the Laguerre function expansion). When truncated to a finite sequence, this method permits numerical calculations of the basic continuum operations often required in applied probability (multiplication, integration, differentiation, and convolution of functions in L_2).

In the developments that follow [11] we expose a "near-closed-form result" for the transient occupancy law on $GI/G/1$ queuing systems (i.e., a result having computational requirement and complexity that are minimal and can be easily implemented using standard commercial or open-source numerical software applications). It is the accuracy and lack of computational burden that makes our approach attractive.

Section 2 is devoted to constructing a computationally useful characterization of the occupancy process for a slight generalization of the $GI/G/1$ queue. Section 3 focuses on a numerical implementation of our results for examining occupancy law dynamics. Section 4 is focused on demonstrating our results of Section 3. We show that for the special case of the $M/M/1$, our approach to constructing occupancy law dynamics for given service and interarrival distributions is no more burdensome than plotting the known closed-form results. With a second example, we demonstrate the insensitivity of our approach to distributional forms, from which we can observe that when given distributions on service times, interarrival times, and batching, we can construct accurate plots of dynamic occupancy law with a computation effort similar to that of plotting the $M/M/1$ closed form.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

2. OCCUPANCY DYNAMICS

All random elements in the developments to follow are defined on the probability space $(\Omega, \mathcal{F}, P)$ described more fully later.

Consider a generalization of the $GI/G/1$ queue for which the following hold:

1. The interarrival times are independent *but not necessarily identically distributed* random variables such that the number of arrivals in $[0, t]$ is finite almost surely, for all $t < \infty$.
2. The service times are independent *but not necessarily identically distributed* random variables.

Each outcome $\omega \in \Omega$ of this queuing system is a sequence of the form $\omega = \{(t_n, s_n); n \geq 1\}$ with nonnegative job arrival times $t_1, t_2, \ldots$ requiring service times $s_1, s_2, \ldots$. Thus, Ω is the set of all sequences of pairs of nonnegative real-valued numbers $\Omega = (\mathbf{R}_+^2)^{(\infty)}$. Let $\mathcal{F}$ be the smallest σ-algebra that includes all finite-dimensional rectangles on Ω. For any n-dimensional rectangle, a probability measure is induced from the sequences of interarrival and service times distributions. Here, P is the unique extension of this probability measure to $\mathcal{F}$.

Define the following random variables:

$T_n,\ n \geq 1$	arrival time of nth job
$S_n,\ n \geq 1$	service time of nth job
$D_n, n \geq 1$	departure time of nth job
$I_n,\ n \geq 1$	interarrival time of nth job ($I_n(\omega) = T_n(\omega) - T_{n-1}(\omega)$; $I_1(\omega) \stackrel{\text{def}}{=} T_1(\omega)$)
$Q_t,\ t \geq 0$	number of jobs in system at time t (occupancy)

Let

$A_n(u) \triangleq P\{I_n \leq u\} \quad \forall u \geq 0,\ n \geq 1$ the interarrival time distribution of nth job,

$B_n(u) \triangleq P\{S_n \leq u\} \quad \forall u \geq 0,\ n \geq 1$ the service time distribution of nth job.

We seek the distribution on the number of jobs in the system at time t:

$$P\{Q_t = k\}, \qquad k \geq 0.$$

Without loss of generality we assume the following:

1. The sample paths for the arrival and departure processes are right-continuous.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

2. There are no jobs in the system prior to time 0 ($Q_{0-} = 0$). Observe that the stochastic process $\{Q_t, t \geq 0\}$, with $Q_{0-} = n$, is equivalent to the process $\{\tilde{Q}_t, t \geq 0\}$ with

$$\begin{aligned} \tilde{Q}_{0-} &= 0, \\ \tilde{A}_i(u) &= 1 \quad \forall u \geq 0, \ \ i = 1, \ldots, n, \\ \tilde{A}_{n+i}(u) &= A_i(u) \quad \forall u \geq 0, \ \ i \geq 1. \end{aligned}$$

2.1. Construction of the Occupancy Process

Fix a time $t \geq 0$. By definition,

$$Q_t = \sum_{i=1}^{\infty} 1_{[0,t]}(T_i) 1_{(t,\infty)}(D_i). \tag{1}$$

Now, define the marked point process $(\mathbf{Y_t}, \mathbf{Z_t}) = \{(Y_{t,n}, Z_{t,n}); n \geq 0\}$ on Ω such that

$$Y_{t,n} = \sum_{i=1}^{n} 1_{[0,t]}(T_i) 1_{(t,\infty)}(D_i), \qquad n \geq 1, \tag{2}$$

$$Y_{t,0} \stackrel{\text{def}}{=} 0,$$

$$Z_{t,n} = 1_{[0,t]}(T_n), \qquad n \geq 1, \tag{3}$$

$$Z_{t,0} \stackrel{\text{def}}{=} 1.$$

It is clear that $Y_{t,n} \to Q_t$ for all $\omega \in \Omega$. Further, $(\mathbf{Y_t}, \mathbf{Z_t})$ enjoys some interesting properties that are useful in our analysis of Q_t. Note that $Y_{t,n}$ represents the number, among the first n arrivals, in the system at time t; $Z_{t,n}$is an indicator function taking the value 1 only if the nth arrival appears by time t.

PROPOSITION 1: *For fixed $t \geq 0$, $(\mathbf{Y_t}, \mathbf{Z_t})$ process forms a Markov chain.*

PROOF: We must show that $(\mathbf{Y_t}, \mathbf{Z_t})$ satisfies the Markov property. For any fixed t and $n \geq 0$,

$$P(Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,i}, Z_{t,i}, i \leq n) = P(Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n}).$$

$Z_{t,n}$ can take only values zero and one. We must demonstrate the Markov property for two cases.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

Case 1: Suppose that $Z_{t,n} = 0$. By definition,

$$Y_{t,n+1} = Y_{t,n} + 1_{[0,t]}(T_{n+1})1_{(t,\infty)}(D_{n+1}) = Y_{t,n} + Z_{t,n+1}1_{(t,\infty)}(D_{n+1}).$$

Hence,

$$\begin{aligned} &P\{Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n} = 0\} \\ &\quad = P\{Y_{t,n} + Z_{t,n+1}1_{(t,\infty)}(D_{n+1}) = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n} = 0\}. \end{aligned}$$

Note that $\{Z_{t,n} = 0\} \subseteq \{T_n > t\} \subseteq \{T_{n+1} > t\} \subseteq \{Z_{t,n+1} = 0\}$. Hence,

$$\begin{aligned} P\{Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n} = 0\} &= P\{Y_{t,n} = y, 0 = z | Y_{t,n}, Z_{t,n} = 0\} \\ &= \begin{cases} 1, & y = Y_{t,n}, z = 0 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

When $Y_{t,n}$ is given, this probability is constant and, therefore, independent of past history. Hence,

$$\begin{aligned} &P\{Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n} = 0\} \\ &\quad = P\{Y_{t,n+1} = y, Z_{t,n+1} = z | Y_{t,n}, Z_{t,n} = 0, Y_{t,i}, Z_{t,i}, i < n\} \end{aligned}$$

and the Markov property holds.

Case 2: Now, suppose that $Z_{t,n} = 1$. We will show that knowledge of $\{Z_{t,n} = 1, Y_{t,n} = k\}$ implies knowledge of the entire history $\{Y_{t,i}, Z_{t,i}; i \le n\}$. Consider the event $\{Y_{t,n} = k, Z_{t,n} = 1\}$. Note that $\{Z_{t,n} = 1\} \subseteq \{T_n \le t\} \subseteq \{T_i \le t, \forall i \le n\} \subseteq \{Z_{t,i} = 1, \forall i \le n\}$. Hence,

$$\{Y_{t,n} = k, Z_{t,n} = 1\} = \{Y_{t,n} = k, Z_{t,i} = 1, \forall i \le n\}.$$

Also, $\{Z_{t,i} = 1, \forall i \le n\}$ implies that

$$\left\{ Y_{t,j} = \sum_{i=1}^{j} Z_{t,i}1_{(t,\infty)}(D_i) \right\} = \left\{ \sum_{i=1}^{j} 1_{(t,\infty)}(D_i) \right\}, \quad \forall j \le n.$$

Noting that $Y_{t,n} = \sum_{i=1}^{n} 1_{(t,\infty)}(D_i) = k$ and since $\{D_i; i \ge 1\}$ is a nondecreasing sequence (jobs are served in order of arrival), it follows that $1_{(t,\infty)}(D_i) = 0, \forall i \le n - k$, and $1_{(t,\infty)}(D_i) = 1, \forall n - k < i \le n$. Hence,

$$\begin{aligned} \{Y_{t,n} = k, Z_{t,n} = 1\} = &\{Y_{t,n} = k, Y_{t,n-1} = k - 1, \ldots, Y_{t,n-k+1} = 1, Y_{t,i} = 0, \\ &\forall i \le n - k, Z_{t,i} = 1, \forall i \le n\} \end{aligned}$$

and the Markov property holds. ■

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

2.2. Arrivals and Departures Adjacent to Time t

Fix a time $t \geq 0$ and define the following random variables on Ω:

$$N_t^a = \inf\{n \geq 1 : T_n > t\}, \tag{4}$$

$$N_t^d = \inf\{n \geq 1 : D_n > t\}, \tag{5}$$

where N_t^a and N_t^d are the indexes of the first arrival and the first departure after time t, respectively. By assumption, the number of arrivals in $[0, t]$ is finite almost surely; therefore, N_t^a is finite almost surely. $T_n > t$ implies that $D_n > t$, so $N_t^d \leq N_t^a$ and N_t^d is also finite almost surely. We note that N_t^a and N_t^d are *stopping times* since $\{N_t^d \leq k\} \in \sigma(\{(Y_{t,i}, Z_{t,i}); i \leq k\})$ and $\{N_t^a \leq k\} \in \sigma(\{(Y_{t,i}, Z_{t,i}); i \leq k\})$. These random variables are useful because when both are given for all t, the entire sample path of the $(\mathbf{Y_t}, \mathbf{Z_t})$ process is revealed.

PROPOSITION 2: *Fix a time $t \geq 0$ with N_t^a and N_t^d defined by* (4) *and* (5)*, respectively. Then*

$$(Y_{t,n}, Z_{t,n}) = \begin{cases} (0, 1), & n < N_t^d \\ (Y_{t,n-1} + 1, 1) = (n - N_t^d + 1, 1), & N_t^d \leq n < N_t^a \\ (Y_{t,n-1}, 0) = (N_t^a - N_t^d, 0), & n \geq N_t^a. \end{cases}$$

PROOF: Fix an $\omega \in \Omega$ and choose an arbitrary index $n \geq 0$.

(a) Case $n < N_t^d$: From (5), $N_t^d > n$ implies that $D_i \leq t$, $\forall i \leq n$. Therefore, from (2), $Y_{t,j} = 0, \forall j \leq n$. Also, $D_n \leq t$ implies that $T_n \leq t$. Thus, $Z_{t,n} = 1_{[0,t]}(T_n) = 1$.

(b) Case $N_t^d \leq n < N_t^a$: From (5), $N_t^d \leq n$ implies that $D_n > t$, and from (4), $N_t^a > n$ implies that $T_n \leq t$. Hence, $1_{[0,t]}(T_n)1_{(t,\infty)}(D_n) = 1$ and $Y_{t,n} = Y_{t,n-1} + 1$. Also, $T_n \leq t$ implies that $Z_{t,n} = 1_{[0,t]}(T_n) = 1$. Finally, $Y_{t,i} = 0, \forall i < N_t^d$, and $Y_{t,i} = Y_{t,i-1} + 1, \forall N_t^d \leq i \leq n$, implies that $Y_{t,n} = n - N_t^d + 1$.

(c) Case $n \geq N_t^a$: Note that $T_n > t$ implies that $1_{[0,t]}(T_n)1_{(t,\infty)}(D_n) = 0$, which, in turn, implies that $Y_{t,n} = Y_{t,n-1}$. Also, $T_n > t$ implies that $Z_{t,n} = 1_{[0,t]}(T_n) = 0$. Finally, $Y_{t,i} = i - N_t^d + 1, \forall N_t^d \leq i < N_t^a$, and $Y_{t,i} = Y_{t,i-1}, \forall i \geq N_t^a$, implies $Y_{t,n} = N_t^a - N_t^d$. ∎

The state transition diagram for $(\mathbf{Y_t}, \mathbf{Z_t})$ is given in Figure 1.

COROLLARY 1:

$$Q_t = \lim_n Y_{t,n} = Y_{t,N_t^a} = N_t^a - N_t^d \quad a.s.$$

PROOF: By definition, N_t^a is finite almost surely (since the number of arrivals in $[0, t]$ is finite a.s.). So, from Proposition 2,

$$\lim_n (Y_{t,n}, Z_{t,n}) = (N_t^a - N_t^d, 0) \quad \text{a.s.}$$

∎

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

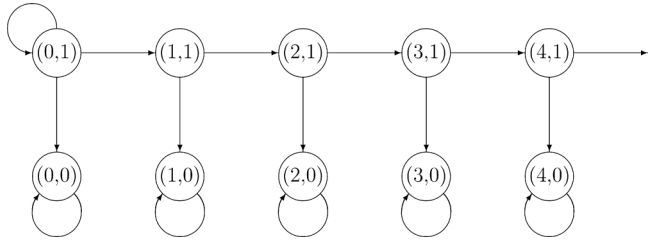


FIGURE 1. State transition diagram of $(\mathbf{Y_t}, \mathbf{Z_t})$.

3. THE OCCUPANCY DISTRIBUTION

PROPOSITION 3: *For any* $t \in \mathbb{R}_+$, *define*

$$q_{t,n,m}(u) = P\{N_t^d = n, T_m \leq u\}, \quad \forall n \geq 1, \forall m \geq n, \forall u \in \mathbf{R}_+. \tag{6}$$

Then

$$P\{Q_t \leq k\} = 1 - \sum_{n=1}^{\infty} q_{t,n,n+k}(t), \quad \forall k \in \mathbf{N_0}.$$

PROOF: Fix an arbitrary nonnegative integer k. From Corollary 1,

$$\begin{aligned} P\{Q_t \leq k\} &= P\{Y_{t,N_t^a} \leq k\} \\ &= P\{N_t^a - N_t^d \leq k\} \\ &= 1 - P\{N_t^a - N_t^d > k\} \\ &= 1 - \sum_{n=1}^{\infty} P\{N_t^a - N_t^d > k, N_t^d = n\} \\ &= 1 - \sum_{n=1}^{\infty} P\{N_t^a > n + k, N_t^d = n\}. \end{aligned}$$

However, from the definition of N_t^a, $\{N_t^a > n + k\} = \{T_{n+k} \leq t\}$. Hence,

$$P\{Q_t \leq k\} = 1 - \sum_{n=1}^{\infty} P\{T_{n+k} \leq t, N_t^d = n\}.$$

Note that the nth term of this sum is $q_{t,n,m}(u)$ (see (6)) for $m = n + k$ and $u = t$. Thus,

$$P\{Q_t \leq k\} = 1 - \sum_{n=1}^{\infty} q_{t,n,n+k}(t). \qquad \blacksquare$$

The usefulness of Proposition 3 in obtaining $P\{Q_t \leq k\}$ is governed by the convergence of the summation. The following proposition addresses this issue. Proposition 3

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

becomes a useful computational result when the sum is truncated to a finite number of terms; the following proposition addresses the precision of such truncations.

PROPOSITION 4: *Fix real numbers* $\epsilon > 0$ *and* $t \geq 0$. *Define*

$$N_\epsilon = \min\{n \in \mathbf{N} : P\{D_n \leq t\} \leq \epsilon\}, \tag{7}$$

$$P_{t,\epsilon}(k) = 1 - \sum_{n=1}^{N_\epsilon} q_{t,n,n+k}(t), \quad \forall k \in \mathbf{N_0}. \tag{8}$$

Then

$$P_{t,\epsilon}(k) - \epsilon \leq P\{Q_t \leq k\} \leq P_{t,\epsilon}(k), \quad \forall k \in \mathbf{N_0}.$$

PROOF: Fix an arbitrary nonnegative integer k. Let N_ϵ be defined by (7). From Proposition 3 it follows that

$$P\{Q_t \leq k\} = 1 - \sum_{n=1}^{N_\epsilon} q_{t,n,n+k}(t) - \sum_{n=N_\epsilon+1}^{\infty} q_{t,n,n+k}(t).$$

Let $P_{t,\epsilon}(k)$ be defined as in (8). Then

$$P\{Q_t \leq k\} = P_{t,\epsilon}(k) - \sum_{n=N_\epsilon+1}^{\infty} q_{t,n,n+k}(t). \tag{9}$$

Clearly, $\sum_{n=N_\epsilon+1}^{\infty} q_{t,n,n+k}\ (t) \geq 0$. Hence,

$$P\{Q_t \leq k\} \leq P_{t,\epsilon}(k) \tag{10}$$

With $q_{t,n,m}(u)$ defined as in (6),

$$\sum_{n=N_\epsilon+1}^{\infty} q_{t,n,n+k}(t) = \sum_{n=N_\epsilon+1}^{\infty} P\{N_t^d = n, T_{n+k} \leq u\} \leq \sum_{n=N_\epsilon+1}^{\infty} P\{N_t^d = n\} = P\{N_t^d > N_\epsilon\}.$$

However, appealing to the definition of N_t^d, we recognize that $P\{N_t^d > N_\epsilon\} = P\{D_{N_\epsilon} \leq t\}$. Further, from the definition of N_ϵ we have that $P\{D_{N_\epsilon} \leq t\} \leq \epsilon$ and, thus, $\sum_{n=N_\epsilon+1}^{\infty} q_{t,n,n+k}(t) \leq \epsilon$. Finally, from (9),

$$P\{Q_t \leq k\} \geq P_{t,\epsilon}(k) - \epsilon.$$

Now, appealing to (10), the proposition follows. ■

Although Proposition 4 offers a result that is amenable to numerical computation, it can be improved. For a fixed k, there can exist a set of indexes n on which $q_{t,n,n+k}$

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

are negligible. Computational speed is improved when such terms are excluded from the sum. The following result formalizes this improvement of Proposition 4.

PROPOSITION 5: *Fix real numbers* $\epsilon > 0$, $t \geq 0$, *and* $0 < r < 1$. *Define*

$$N_{\epsilon/2} = \min\{n \geq 1 : P\{D_n \leq t\} \leq \epsilon/2\}, \tag{11}$$

$$\mathbf{E}_{\epsilon,r} = \{n \geq 1 : n \leq N_{\epsilon/2}, q_{t,n,n+k}(t) > (1-r)r^{n-1}\epsilon/2\}, \tag{12}$$

and

$$P_{t,\epsilon,r}(k) = 1 - \sum_{n \in \mathbf{E}_{\epsilon,r}} q_{t,n,n+k}(t), \quad \forall k \in \mathbf{N_0} \tag{13}$$

Then

$$P_{t,\epsilon,r}(k) - \epsilon \leq P\{Q_t \leq k\} \leq P_{t,\epsilon,r}(k), \quad \forall k \in \mathbf{N_0}.$$

PROOF: Fix an arbitrary nonnegative integer k. From (11) and Proposition 3,

$$P_{k,\epsilon/2} - \epsilon/2 \leq P\{Q_t \leq k\} \leq P_{k,\epsilon/2}, \tag{14}$$

where

$$P_{k,\epsilon/2} = 1 - \sum_{n=1}^{N_{\epsilon/2}} q_{t,n,n+k}(t).$$

Let us partition the index set of the sum in two disjoint subsets

$$\{n \geq 1 : n \leq N_{\epsilon/2}\} = \mathbf{E}_{\epsilon,r} \cup \mathbf{E}^c_{\epsilon,r},$$

where $\mathbf{E}_{\epsilon,r}$ is defined by (12) and

$$\begin{aligned}\mathbf{E}^c_{\epsilon,r} &= \{n \geq 1 : n \leq N_{\epsilon/2}\} \backslash \mathbf{E}_{\epsilon,r} \\ &= \{n \geq 1 : n \leq N_{\epsilon/2}, q_{t,n,n+k}(t) \leq (1-r)r^{n-1}\epsilon/2\}.\end{aligned}$$

Then

$$P_{k,\epsilon/2} = 1 - \sum_{n \in \mathbf{E}_{\epsilon,r}} q_{t,n,n+k}(t) - \sum_{n \in \mathbf{E}^c_{\epsilon,r}} q_{t,n,n+k}(t),$$

and the first two terms on right correspond to the definition of $P_{t,\epsilon,r}(k)$ (see (13)). Therefore,

$$P_{k,\epsilon/2} = P_{t,\epsilon,r}(k) - \sum_{n \in \mathbf{E}^c_{\epsilon,r}} q_{t,n,n+k}(t). \tag{15}$$

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

However, $q_{t,n,n+k}(t) \leq (1-r)r^{n-1}\epsilon/2$ on $\mathbf{E}^c_{\epsilon,r}$. Substituting for $q_{t,n,n+k}(t)$ its maximum value and extending the sum for every positive integer,

$$\begin{aligned} P_{k,\epsilon/2} &\geq P_{t,\epsilon,r}(k) - \sum_{\{n\geq 1\}} (1-r)r^{n-1}\epsilon/2 \\ &\geq P_{t,\epsilon,r}(k) - (1-r)\epsilon/2 \sum_{n=1}^{\infty} r^{n-1} \\ &\geq P_{t,\epsilon,r}(k) - \epsilon/2. \end{aligned}$$

Thus,

$$\begin{aligned} P_{t,\epsilon,r}(k) &\leq P_{k,\epsilon/2} + \epsilon/2, \\ P_{t,\epsilon,r}(k) - \epsilon &\leq P_{k,\epsilon/2} - \epsilon/2. \end{aligned}$$

From (14),

$$P_{t,\epsilon,r}(k) - \epsilon \leq P\{Q_t \leq k\}. \tag{16}$$

Now, from (15) it is clear that $P_{k,\epsilon/2} \leq P_{t,\epsilon,r}(k)$, and then from (14),

$$P\{Q_t \leq k\} \leq P_{t,\epsilon,r}(k),$$

which completes the proof. ■

Proposition 5 suggests a methodology for obtaining ϵ-bounds on the time-dependent occupancy distribution. It remains to compute $q_{t,n,m}(t)$; in order to accomplish this, we must introduce some auxiliary distributions.

For clarity in the development of our arguments to follow, we employ a slight abuse of notation (similar to that used by Taylor and Karlin [14]), writing conditional distributions (appearing within integrand expressions) as conditional probabilities.

Let $J_{m,n}(u,v)$ denote the joint distribution of the mth arrival and the nth departure times; that is,

$$J_{m,n}(u,v) = P\{T_m \leq u, D_n \leq v\}, \qquad m,n \geq 1, \tag{17}$$

$$J_{1,0}(u,v) = P\{T_1 \leq u\} = A_1(u). \tag{18}$$

Proposition 6: *Let $q_{t,n,m}(u)$ be defined by* (6) *and $J_{m,n}(u,v)$ be defined by* (17) *and* (18). *Then*

$$q_{t,n,n}(u) = J_{n,n-1}(u,t) - J_{n,n}(u,t), \qquad n \geq 1, \tag{19}$$

$$q_{t,n,m}(u) = q_{t,n,m-1} * A_m(u), \qquad n \geq 1, m > n, \tag{20}$$

where $*$ is the convolution operator.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

PROOF:

(a) With $n = m = 1$, it follows from (6) that

$$\begin{aligned} q_{t,1,1}(u) &= P\{T_1 \le u, N_t^d = 1\} \\ &= P\{T_1 \le u, D_1 > t\} \\ &= P\{T_1 \le u\} - P\{T_1 \le u, D_1 \le t\} \\ &= A(u) - J_{1,1}(u,t). \end{aligned}$$

However, from (18), $J_{1,0}(u,t) = A(u)$; hence,

$$q_{t,1,1}(u) = J_{1,0}(u,t) - J_{1,1}(u,t)$$

and (19) holds.

(b) With $n = m$, $n > 1$, it follows from (6) that

$$\begin{aligned} q_{t,n,n}(u) &= P\{T_n \le u, N_t^d = n\} \\ &= P\{T_n \le u, D_{n-1} \le t, D_n > t\} \\ &= P\{T_n \le u, D_{n-1} \le t\} - P\{T_n \le u, D_{n-1} \le t, D_n \le t\} \\ &= P\{T_n \le u, D_{n-1} \le t\} - P\{T_n \le u, D_n \le t\}. \end{aligned}$$

Then, from (17),

$$q_{t,n,n}(u) = J_{n,n-1}(u,t) - J_{n,n}(u,t).$$

(c) With $m > n$,

$$\begin{aligned} q_{t,n,m}(u) &= P\{T_m \le u, N_t^d = n\} \\ &= P\{T_{m-1} + I_m \le u, N_t^d = n\} \\ &= \int_0^\infty P\{T_{m-1} + I_m \le u, N_t^d = n | I_m = x\}\, dP\{I_m \le x\} \\ &= \int_0^\infty P\{T_{m-1} \le u - x, N_t^d = n | I_m = x\}\, dA_m(x). \end{aligned}$$

Clearly, the integrand is zero for $x > u$. Also, from definition (5), $\{N_t^d = n\} = \{D_{n-1} \le t, D_n > t\}$; thus,

$$q_{t,n,m}(u) = \int_0^u P\{T_{m-1} \le u - x, D_{n-1} \le t, D_n > t | I_m = x\}\, dA_m(x).$$

However, I_m is independent of T_{m-1} and is independent of $D_n, \forall n < m$; therefore,

$$q_{t,n,m}(u) = \int_0^u P\{T_{m-1} \le u - x, D_{n-1} \le t, D_n > t\}\, dA_m(x).$$

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

Once again, noting that $\{N_t^d = n\} = \{D_{n-1} \leq t, D_n > t\}$,

$$q_{t,n,m}(u) = \int_0^u P\{T_{m-1} \leq u - x, N_t^d = n\}\, dA_m(x).$$

From definition (6),

$$\begin{aligned} q_{t,n,m}(u) &= \int_0^u q_{t,n,m-1}(u - x)\, dA_m(x) \\ &= q_{t,n,m-1} * A_m(u). \end{aligned}$$

■

Finally, we must obtain $J_{n,n}$ and $J_{n+1,n}$ for all $n \geq 1$. Proposition 7 suggests an iterative approach for obtaining these distributions.

PROPOSITION 7: *Let $J_{m,n}$ be defined by* (17) *and* (18). *Then, for all $n \geq 1$,*

$$J_{n,n}(u, v) = \int_{[0,v]} J_{n,n-1}(\min(u, v - y), v - y)\, dB_n(y), \tag{21}$$

$$J_{n+1,n}(u, v) = \int_{[0,u]} J_{n,n}(u - x, v)\, dA_{n+1}(x). \tag{22}$$

Defining the functions $C_v : \mathbf{R} \to \mathbf{R}^2$ and $R_u : \mathbf{R} \to \mathbf{R}^2$ as

$$C_v(u) = (u, v), \quad \forall v \geq 0, u \geq 0, \tag{23}$$

$$R_u(v) = (\min(u, v), v), \quad \forall u \geq 0, v \geq 0, \tag{24}$$

(21) *and* (22) *can be expressed as:*

$$J_{n,n}(u, v) = (J_{n,n-1} \circ R_u) * B_n(v), \tag{25}$$

$$J_{n+1,n}(u, v) = (J_{n,n} \circ C_v) * A_{n+1}(u), \tag{26}$$

where $\circ$ represents the composition operation.

PROOF:

(a) First, we will show that (21) is true for the initial condition ($n = 1$). By definition,

$$J_{1,1}(u, v) = P\{T_1 \leq u, D_1 \leq v\}.$$

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

By assumption, there are no jobs in the system prior to time 0; therefore, $D_1 = T_1 + S_1$ and

$$\begin{aligned} J_{1,1}(u,v) &= P\{T_1 \le u, T_1 + S_1 \le v\} \\ &= \int_{[0,\infty)} P\{T_1 \le u, T_1 + S_1 \le v | S_1 = y\}\, dP\{S_1 \le y\} \\ &= \int_{[0,\infty)} P\{T_1 \le u, T_1 \le v - y | S_1 = y\}\, dB_1(y). \end{aligned}$$

However, clearly this integrand is zero for $y > v$. Thus,

$$J_{1,1}(u,v) = \int_{[0,v]} P\{T_1 \le u, T_1 \le v - y | S_1 = y\}\, dB_1(y).$$

By assumption, S_1 is independent of T_1 and

$$\begin{aligned} J_{1,1}(u,v) &= \int_{[0,v]} P\{T_1 \le \min(u, v-y)\}\, dB_1(y) \\ &= \int_{[0,v]} A_1(\min(u, v-y))\, dB_1(y). \end{aligned} \tag{27}$$

By definition, $J_{1,0}(u,v) = A_1(u)$ and

$$J_{1,1}(u,v) = \int_{[0,v]} A_1(\min(u, v-y))\, dB_1(y)$$

holds for $n = 1$.

(b) For the case of $n > 1$, suppose that $J_{n,n-1}(u,v)$ is known. Then

$$\begin{aligned} J_{n,n}(u,v) &= P\{T_n \le u, D_n \le v\} \\ &= P\{T_n \le u, \max(T_n, D_{n-1}) + S_n \le v\} \\ &= P\{T_n \le u, T_n + S_n \le v, D_{n-1} + S_n \le v\} \\ &= \int_{[0,\infty)} P\{T_n \le u, T_n + S_n \le v, D_{n-1} + S_n \le v | S_n = y\}\, dP\{S_n \le y\} \\ &= \int_{[0,\infty)} P\{T_n \le u, T_n \le v - y, D_{n-1} \le v - y | S_n = y\}\, dB_n(y). \end{aligned}$$

However, clearly, this integrand is zero for $y > v$ and

$$J_{n,n}(u,v) = \int_{[0,v]} P\{T_n \le u, T_n \le v - y, D_{n-1} \le v - y | S_n = y\}\, dB_n(y).$$

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

By assumption, S_n is independent of T_n and D_{n-1}; thus,

$$J_{n,n}(u,v) = \int_{[0,v]} P\{T_n \le \min(u, v-y), D_{n-1} \le v-y\}\, dB_n(y)$$
$$= \int_{[0,v]} J_{n,n-1}(\min(u, v-y), v-y)\, dB_n(y).$$

Hence, (21) holds in general. In order to obtain the equivalent expression (25), define $R_u : \mathbf{R} \to \mathbf{R}^2$ by (24); then

$$J_{n,n}(u,v) = \int_{[0,v]} J_{n,n-1} \circ R_u(v-y)\, dB_n(y)$$
$$= (J_{n,n-1} \circ R_u) * B_n(v).$$

(c) Now, suppose that $J_{n,n}(u,v), n \ge 1$, is given. Then

$$J_{n+1,n}(u,v) = P\{T_{n+1} \le u, D_n \le v\}$$
$$= P\{T_n + S_{n+1} \le u, D_n \le v\}$$
$$= \int_{[0,\infty)} P\{T_n + I_{n+1} \le u, D_n \le v | I_{n+1} = x\}\, dP\{I_{n+1} \le x\}$$
$$= \int_{[0,\infty)} P\{T_n \le u-x, D_n \le v | I_{n+1} = x\}\, dA_{n+1}(x).$$

Clearly, the integrand is zero for $x > u$. Thus,

$$J_{n+1,n}(u,v) = \int_{[0,u]} P\{T_n \le u-x, D_n \le v | I_{n+1} = x\}\, dA_{n+1}(x).$$

By assumption, I_{n+1} is independent of T_n and D_n; therefore,

$$J_{n+1,n}(u,v) = \int_{[0,u]} P\{T_n \le u-x, D_n \le v\}\, dA_{n+1}(x)$$
$$= \int_{[0,u]} J_{n,n}(u-x, v)\, dA_{n+1}(x)$$

and (22) holds. In order to prove (26), define $C_v : \mathbf{R} \to \mathbf{R}^2$ by (23); it then follows that

$$J_{n+1,n}(u,v) = \int_{[0,u]} J_{n,n} \circ C_v(u-x)\, dA_{n+1}(x)$$
$$= (J_{n,n} \circ C_v) * A_{n+1}(u).$$

■

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

From Propositions 5–7, we can obtain ϵ-bounds $(P_{t,\epsilon,r}(k))$ for $P(Q_t \leq k), \forall k \geq 0$. The algorithm of Figure 2 formalizes such a computational approach, yielding $P_{t,\epsilon,r}(k)$ for all $k \geq 0$. It is important to note that Proposition 5 suggests fixing a value k and calculating $q_{t,n,n+k}(t)$ for all $n \geq 1$ in order to obtain $P_{t,\epsilon,r}(k)$. The approach represented by the pseudocode shown in Figure 2 is different. Here, we fix a value n and calculate $q_{t,n,n+k}(t)$ for all k in the index set $E_{\epsilon,r}$ defined by (12), then continue with the next n, and so forth. In other words, all $P_{t,\epsilon,r}(k)$ are computed simultaneously, which is better suited for numerical computation. The pseudocode reveals that the computational burden of our approach is insensitivity to the form of input distribution functions.

No.	Algorithm step	Justification
1	$P_{t,\epsilon,r}(k) = 1; \forall k \geq 0$	Proposition 5
2	$n = 0$	
3	repeat	
4	$\quad n = n + 1$	
5	$\quad J_{n,n-1}(u,v) = \begin{cases} A_1(u) & n = 1 \\ (J_{n-1,n-1} \circ C_v) * A_n(u) & n > 1 \end{cases}$	Proposition 7
6	$\quad J_{n,n}(u,v) = (J_{n,n-1} \circ R_u) * B_n(v)$	Proposition 7
7	$\quad q_{t,n,n}(u) = J_{n,n-1}(u,t) - J_{n,n}(u,t)$	Proposition 6
8	$\quad k = 0$	
9	$\quad P_{t,\epsilon,r}(k) = P_{t,\epsilon,r}(k) - q_{t,n,n+k}(t)$	Proposition 5
10	$\quad$ repeat	
11	$\quad\quad k = k + 1$	
12	$\quad\quad q_{t,n,n+k}(u) = q_{t,n,n+k-1} * A_{n+k}(u)$	Proposition 6
13	$\quad\quad P_{t,\epsilon,r}(k) = P_{t,\epsilon,r}(k) - q_{t,n,n+k}(t)$	Proposition 5
14	$\quad$ until $q_{t,n,n+k}(t) \leq (1-r)\, r^{n-1}\epsilon/2$	Proposition 5
15	until $J_{n,n}(t,t) \leq \epsilon/2$	Proposition 5
		$(J_{n,n}(t,t) = P(D_n \leq t))$

FIGURE 2. Pseudocode to compute $P_{t,\epsilon,r}(k)$.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

4. EXAMPLES

It is useful to demonstrate our approach to capturing the time-dependent law on system occupancy through a pair of straightforward examples. To this end, we constructed an implementation of the pseudocode shown in Figure 2 in C + +. Our purpose is not to develop the most computationally efficient code; rather, we seek to demonstrate the correctness and accuracy of our analytical developments.

For numerical integration, we employ the well-known composite Simpson's rule [3, Sect. 4.3], with a slight modification. This rule requires an even number of integration domain partition elements; in the cases where the cardinality is odd, Simpson's three-eighths rule is used for the first three subintervals and the standard Simpson's rule elsewhere. We will refer to this approach as a *modified composite Simpson's rule* (MCSR).

4.1. Comparison with Well–Known $M/M/1$ Results

For the purpose of verifying the correctness of our analytical approach, we will examine the only special case of a $G/G/1$ queuing system for which there exists a closed-form expression for the time-dependent probability law on occupancy: the $M/M/1$ queue. This result is widely available and cited, for example, by Bailey [7, Sect. 2.10.2], where it is reported that

$$P_t(n) = e^{-(\lambda+\mu)t}\left[\left(\frac{\mu}{\lambda}\right)^{(i-n)/2} I_{n-i}\left(2\sqrt{\lambda\mu}\,t\right) + \left(\frac{\mu}{\lambda}\right)^{(i-n+1)/2} I_{n+i+1}\left(2\sqrt{\lambda\mu}\,t\right) + \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n} \sum_{l=n+i+2}^{\infty} \left(\frac{\mu}{\lambda}\right)^{(l/2)} I_l\left(2\sqrt{\lambda\mu}\,t\right)\right], \qquad n \geq 1, \tag{28}$$

with $P_t(n) = P\{Q_t = n\}$, λ is the arrival rate, μ is the service rate, i is the number of jobs in system at time 0, and $I_n(y)$ is the modified Bessel function of the first kind.

We note that for the purposes of numerical computations, the closed-form expression of (28) is not necessarily simple to use, since it requires evaluating a weighted infinite sum of modified Bessel functions of the first kind. Further, the boundary condition $P_t(0)$ is characterized using yet another infinite sum:

$$P_t(0) = 1 - \sum_{n=1}^{\infty} P_t(n). \tag{29}$$

We employ *Scientific Notebook* to obtain values from (28) against which we could draw a comparison. This software incorporates the highly touted Maple™ kernel for numerical calculations, which includes embedded routines for modified Bessel functions of the first kind.

Thus, we provide the function $P_t(n)$ (28) to the package, specifying numerical values for λ, μ, i, and t, and *Scientific Notebook* automatically returns a numerical

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

value having a claimed precision of 10 decimals places. For the boundary constraint of (29), we employ the truncated series

$$P_t(0) = 1 - \sum_{n=1}^{N} P_t(n),$$

where N is assigned a large integer. For this verification check, we specify that the system is initially empty and idle ($i = 0$), with a traffic intensity of 0.8. Without loss of generality, we take $\lambda = 1$; thus, $\mu = 1/0.8 = 1.25$.

Scientific Notebook could compute values only up to $t = 16$. For larger time values, *Scientific Notebook* was (not surprisingly) unable to appropriately approximate the required infinite sums. Hence, we report only results for $t \leq 10$.

The numerical values returned by *Scientific Notebook*, when evaluating the closed-form expression of the occupancy distribution for $t \leq 10$, were compared with results obtained using our methodology and code. Our code was run to obtain $P_{t,\epsilon,r}(k)$ for $t = 10, \epsilon = 0.0001, r = 0.9$, and $N = 150$, yielding a quadrature rule error estimate that was less than ϵ. With $t \leq 10$, computed at increments of 0.05, a plot of the time-dependent occupancy distribution is shown in Figure 3.

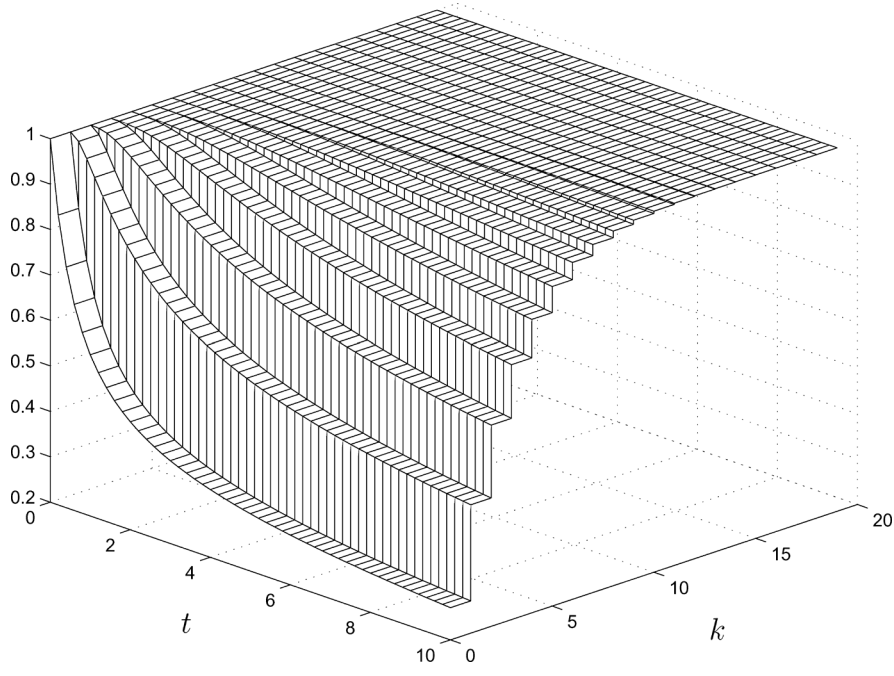


FIGURE 3. [$\widehat{P}_{t,0.0001,0.9}(k)$ obtained with MCSR for $N = 200$], tightly approximating $P\{Q_t \leq k\}$.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

Our tight approximation $\widehat{P}_{t,\epsilon,r}(k)$ of $P\{Q_t \leq k\}$ yielded a maximum absolute error deviation of 0.0000214 from the values obtained from the closed-form expression using the *Scientific Notebook* and the embedded kernel.[1] Again, our purpose with this example is simply to support verification of our analytical approach; it is not our purpose to claim the efficiency of our computational code. We note that our implementation executes so quickly (well under 1 min on a desktop environment) that the computation time is, for most practical purposes, insignificant.

4.2. Time-Dependent Occupancy Distribution for a $Gamma^{[x]}/Weibull/1$ Queuing System

As previously noted, our analytical approach allows analyses of a modest generalization of the $G/G/1$ queuing system in that we can accommodate interarrival times and service times that, although mutually independent, are not necessarily identically distributed. We demonstrate this by constructing the time-dependent occupancy distribution for the following scenario.

Consider a queuing system in which jobs arrive in batches of size 3. Jobs are served one at time, and each job departs the system upon service completion; we seek the time-dependent system occupancy. Assume that the interarrival times are independent and follow the distribution Gamma($\alpha = 2, \beta = 5$) and that job service times follow Weibull($\alpha = 2, \beta = 3$).

Since the Gamma and Weibull distributions belongs to $C^{\infty}(0, t]$ for any real number $t \geq 0$ and their density functions are known, MCSR can be used. The interarrival distributions and the corresponding densities are, respectively, given by

$$A_n(t) = \begin{cases} 1 - (1 + t/5)e^{-t/5} & n = 1, 4, 7, 10, \ldots \\ 1 & \text{otherwise,} \end{cases}$$

$$a_n(t) = \begin{cases} \dfrac{1}{25} te^{-t/5} & n = 1, 4, 7, 10, \ldots \\ \delta(t) & \text{otherwise,} \end{cases}$$

where $\delta(t)$ is the Dirac delta function. The service time density function is

$$b_n(t) = \frac{2}{9} te^{-t^2/9}, \qquad n =\geq 1.$$

With these functions, our methodology and code were executed using the MCSR as a quadrature rule. N was selected in such a way that the maximum quadrature

[1] Not surprisingly, the computational effort required by our approach to construct this region of the time-dependent occupancy distribution for the $M/M/1$ example is an order of magnitude faster than construction via Maple.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

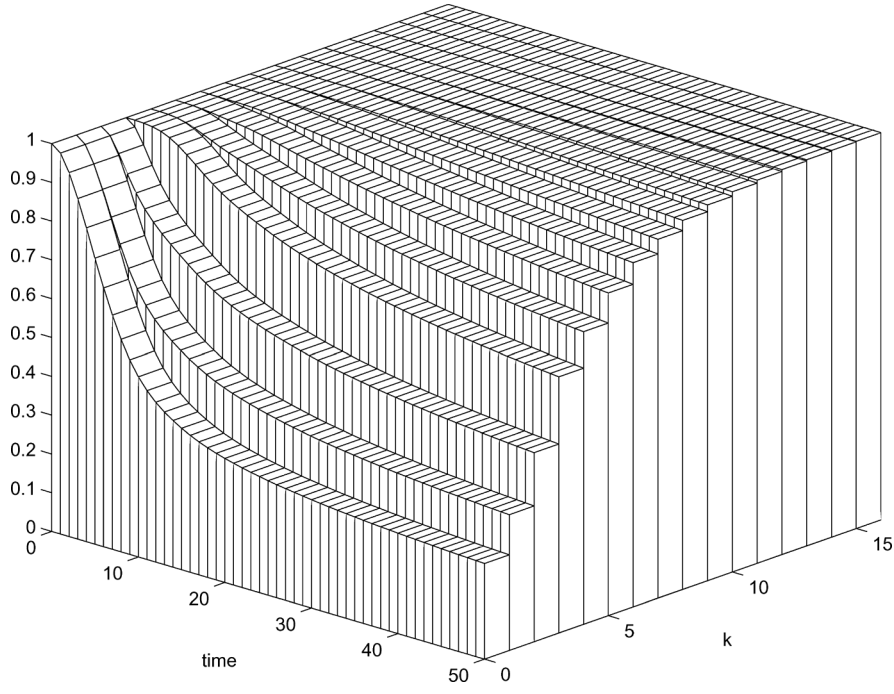


FIGURE 4. $\widehat{P}^{MCSR,200}_{t,0.0001,0.9}(k)$ for the system $Gamma^{[3]}(2,5)/Weibull(2,3)/1$.

error was less than 0.0001, leading to $N = 200$. A maximum value of $t = 50$ was selected so that the transient region of the distribution appears clearly in the plot. The time-dependent occupancy distribution was constructed for each time increment, with $\epsilon = 0.0001$ and $r = 0.9$.

The value of $\widehat{P}^{MCSR,200}_{t,0.0001,0.9}(k)$ is shown in Figure 4. The estimated quadrature error was computed for all points in the mesh shown in this figure. The maximum estimated quadrature error was 1.708×10^{-5} (at the point $t = 50, k = 15$). Given the resulting small quadrature with $\epsilon = 0.0001$, Figure 4 represents an accurate approximation of $P\{Q_t \leq k\}$. Again, an execution time of well under 1 min was achieved on a desktop workstation.

Here, we note that although a closed-form transient law on the $GI/G/1$ queue is unavailable, a straightforward numerical procedure allows one to plot the transient law of (a generalization of) the $GI/G/1$ with no more effort than plotting the closed-form transient law on the $M/M/1$. We believe that such "near-closed-form" results merit close investigation for generalizations beyond the queuing paradigms considered here.

Acknowledgment

This work was supported by National Science Foundation grant DMI–0233669.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press

References

1. Agrawala, A.K. & Tripathi, S.K. (1980). Transient solution of the virtual waiting time of a single-server queue and its applications. *Information Sciences* 21: 141–158.
2. Borovkov, A. (1976). *Stochastic processes in queueing theory*. Berlin: Springer-Verlag.
3. Burden, R.L. & Faires, J.D. (1993). *Numerical analysis*, 5th ed. Boston: PWS Publishing.
4. Chaudhry, M., Agarwal, M., & Templeton, J. (1992). Exact and approximate numerical solutions of steady-state distributions arising in the queue GI/G/1. *Queueing Systems* 10: 105–152.
5. Choudhury, G.L., Lucantoni, D.M., & Whitt, W. (1997). Numerical solution of the piecewise-stationary $M_t/G_t/1$ queues. *Operations Research* 45(3): 451–463.
6. Cohen, J. (1982). *The single server queue*, 2nd ed. New York: Elsevier Science Publishing.
7. Gross, D. & Harris, C.M. (1985). *Fundamentals of queueing theory*, 2nd ed. New York: Wiley.
8. Keilson, J. & Nunn, W. (1979). Laguerre transformation as a tool for the numerical solution of integral equations of convolution type. *Applied Mathematics and Computation* 5: 313–359.
9. Kingman, J. (1961). The single server queue in heavy traffic. *Proceedings of the Cambridge Philospohical Society* 57: 902–904.
10. Kingman, J. (1962). Some inequalities for the queue GI/G/1. *Biometrika* 49(3/4): 315–324.
11. Limón-Robles, J. (1998). On the time dependent occupancy distribution of the $G/G/1$ queueing system. Ph.D. dissertation, Monterrey, Mexico: Instituto Tecnológico y de Estudios Superiores de Monterrey.
12. Litko, J.R. (1989). GI/G/1 interdeparture time and queue-length distributions via the Laguerre transform. *Queueing Systems* 4: 367–382.
13. Marchal, W.G. (1978). Some simpler bounds on the mean queueing time. *Operations Research* 26(6): 1083–1088.
14. Taylor, H. & Karlin, S. (1998). *Introduction to stochastic modeling*, 3rd ed. San Diego, CA: Academic Press.
15. Whitt, W. (1993). Approximations for the $GI/G/m$ queue. *Production and Operations Management* 2(2): 114–161.
16. Yang, T. & Chaudhry, M. (1996). On steady-state queue size distributions of the discrete-time GI/G/1 queue. *Advances in Applied Probability* 28: 1177–1200.

https://doi.org/10.1017/S0269964809000163 Published online by Cambridge University Press