The Gated, Infinite-Server Queue: Uniform Service Times

Sid Browne[1], *E. G. Coffman, Jr.*[2], *E. N. Gilbert*[2], *Paul E. Wright*[2]

[1] Columbia University, New York, NY 10027,
[2] AT&T Bell Laboratories, Murray Hill, NJ 07974

ABSTRACT

Customers, in a Poisson stream at rate λ, enter an infinite-server queue. Customer service times are independent and uniformly distributed on $[0, s]$, $s > 0$. Gated service is performed in stages as follows. A stage begins with all customers transferred from the queue to the servers. The servers then begin serving these customers, all simultaneously. The stage ends when the service of all customers is complete. At this point, the next stage begins if the queue is nonempty. If the queue is empty, the servers just remain idle awaiting the next arrival, at which time the next stage begins.

This paper develops asymptotics for the equilibrium distribution of the number served in a stage in the light and heavy traffic regimes $\lambda \to 0$ and $\lambda \to \infty$, respectively. The results are obtained by the analysis of a Fredholm integral equation of the second kind. For computational purposes, the integral equation is transformed into an infinite system of linear algebraic equations. The effect of truncating the system to a finite size is then examined.

The Gated, Infinite-Server Queue: Uniform Service Times

Sid Browne[1], *E. G. Coffman, Jr.*[2], *E. N. Gilbert* [2], *Paul E. Wright* [2]

[1] Columbia University, New York, NY 10027,
[2] AT&T Bell Laboratories, Murray Hill, NJ 07974

1 Introduction

Customers, in a Poisson stream at rate λ, enter an infinite-server queue. Service is gated in that customers are served in stages, as follows. A stage begins with all customers transferred from the queue to the servers (the gate opens to admit waiting customers and then closes). The servers then begin serving these customers, all simultaneously. The stage ends when the service of all customers is complete. At this point, the next stage begins if the queue is nonempty. If the queue is empty, the servers just remain idle awaiting the next arrival, at which time the next stage begins.

Let k_n denote the number of requests served during the n^{th} stage. Customer service times are assumed to be independent, so $\{k_n\}$ is easily seen to be a Markov chain. This paper analyzes the behavior of $\{k_n\}$ when service times are uniformly distributed on $[0, s]$, $s > 0$. The normalization $s = 1$ is convenient and applies hereafter. By the analysis of a Fredholm integral equation of the second kind, asymptotics are developed for the mean, variance and generating function of the equilibrium distribution in the light and heavy traffic regimes. In light traffic, when λ is small, almost all stages serve only one customer. In heavy traffic, when λ is large, stages serve many customers, so a stage duration (the maximum of the service times in the stage) is nearly one time unit. Then k_n is approximately Poisson distributed with mean λ. This paper derives asymptotic series that give higher order corrections to these approximations. For other work on asymptotic solutions of Fredholm equations of the type considered here, see [5, 6, 8].

A considerable literature exists on the analysis of gating disciplines restricted to single-server systems. Recent examples are [1, 4, 7]. In [2], a gated infinite-server queue with vacations was studied. The infinite-server system "goes on vacation" whenever the server finds the queue empty on its return. While the model here is the special case without vacations and with uniformly distributed service times, the present analysis of $\{k_n\}$ is carried to greater depth.

As noted in [2], there are many applications modeled by gated, parallel service. Among these are data transmission stations, in which servers are communication channels, and task-oriented parallel simulations, in which tasks are repeatedly removed from a queue by servers (processors) working in parallel. In general terms, the gated, infinite-server queue applies to those systems in which a large number of parallel servers require a gating mechanism to synchronize the servers following each stage of a computation. For parallel simulations the synchronization points are often necessary to guarantee the correctness of computations. The reader will find further discussion in [2].

Acknowledgement It is a pleasure to thank J. A. Morrison for his gracious assistance with the heavy traffic analysis of Section 4.

2 Preliminaries

This section begins with a brief reprise of the results established in [2] that are required for the present analysis. For the Markov chain $\{k_n\}$, let $p_k^{(n)} = \Pr\{k_n = k\}$, $k \geq 1$, and define the generating function

$$P^{(n)}(y) = \sum_{k=1}^{\infty} p_k^{(n)} y^k \ .$$

Let $H^{(n)}(t)$ denote the probability distribution of the n^{th} stage duration. The $(n+1)^{\text{st}}$ stage serves 1 customer if there were 0 arrivals or 1 arrival during the n^{th} stage. But $k > 1$ customers are served in the $(n+1)^{\text{st}}$ stage if k customers arrived during the n^{th} stage. Then

$$\begin{aligned} p_1^{(n+1)} &= \int_0^1 [e^{-\lambda t} + \lambda t e^{-\lambda t}] dH^{(n)}(t) \\ p_k^{(n+1)} &= \int_0^1 \frac{(\lambda t)^k}{k!} e^{-\lambda t} dH^{(n)}(t) \ . \end{aligned}$$

To determine the generating function for this distribution, note that, by the uniform distribution of service times, t^k is the probability distribution of the duration of a stage serving k customers. Then an easy calculation gives for the function $Q^{(n)}(y) = 1 - P^{(n)}(y)$,

$$Q^{(n+1)}(y) \;=\; \lambda(1-y)\int_0^1 [e^{-\lambda(1-y)t} - e^{-\lambda t}]Q^{(n)}(t)dt + 1 - y \;. \tag{2.1}$$

This recurrence for $Q^{(n)}(y)$ may be abbreviated to

$$Q^{(n+1)}(y) \;=\; \mathbf{K}Q^{(n)}(y) + 1 - y \;,$$

where $\mathbf{K}$ is the integral operator

$$\mathbf{K}f(y) \;=\; \int_0^1 K(y,t)f(t)dt \;,$$

with the kernel

$$K(y,t) \;=\; \lambda(1-y)[e^{-\lambda(1-y)t} - e^{-\lambda t}] \;. \tag{2.2}$$

Iterating (2.1) gives the solution

$$Q^{(n)}(y) \;=\; \mathbf{K}^n Q^{(0)}(y) + (\mathbf{I} + \mathbf{K} + \cdots + \mathbf{K}^{n-1})(1-y) \;. \tag{2.3}$$

In [2] it was shown that $\mathbf{K}^n Q^{(0)}(y) \to 0$ as $n \to \infty$ for all $Q^{(0)}(y)$, while the remaining terms of (2.3) approach a convergent series

$$Q(y) \;=\; (\mathbf{I} + \mathbf{K} + \mathbf{K}^2 + \cdots)(1-y) \;. \tag{2.4}$$

For any initial state distribution, described by $Q^{(0)}(y)$, (2.3) represents a transient towards a stationary distribution, described by $Q(y)$. Our objective is to find $Q(y)$ in order to obtain the stationary probabilities $p_k = \lim_{k\to\infty} p_k^{(n)}$ from $\sum p_k y^k = P(y) = 1 - Q(y)$.

Equation (2.4) may be recognized as the Neumann series for solving the integral equation

$$Q(y) \;=\; \mathbf{K}Q(y) + 1 - y \tag{2.5}$$

(see [3, 9]). Partial sums of (2.4) are the functions $Q^{(n)}(y)$ in (2.3) that are obtained from the iteration (2.1), starting with $Q^{(0)}(y) = 1 - y$. With that initial distribution, $Q^{(n)}(y)$ describes the queue n stages after an initial stage that served a single customer (such as a stage following an idle period).

Sections that follow will derive good approximations to $Q(y)$. These approximations will then supply good approximations to the state probabilities p_k. For, after expanding the kernel $K(y,t)$ in a power series in y, one may equate coefficients in (2.5) and find

$$p_k \;=\; \begin{cases} 1 - \lambda^2 \int_0^1 te^{-\lambda t} Q(t)dt, & k = 1 \\ \lambda \int_0^1 \left(\frac{(\lambda t)^{k-1}}{(k-1)!} - \frac{(\lambda t)^k}{k!} \right) e^{-\lambda t} Q(t) dt, & k \geq 2 \; . \end{cases} \tag{2.6}$$

Numerical approximations to $Q(y)$ also have interest because of the connection between $Q(y)$ and the duration of a stage, i.e., $1 - Q(y) = \sum p_k t^k$ is the stationary distribution of stage durations.

Intuition may suggest that the Markov chain approaches stationarity rapidly, especially if λ is small or large. That entails rapid convergence of the Neumann series (2.4). One bound on the rate of convergence of (2.4) is obtained from the norm

$$||\mathbf{K}|| \;=\; \sup_{f,y} |\mathbf{K}f(y)| \; , \tag{2.7}$$

in which $0 \leq y \leq 1$ and f is restricted to integrable functions with $\sup_{0 \leq y \leq 1} |f(y)| = 1$. If $||\mathbf{K}|| < 1$, (2.4) will converge at least as fast as the power series for $1/(1 - ||\mathbf{K}||)$. Let

$$\mathbf{K}1(y) = \int_0^1 K(y,t)dt = 1 - e^{-\lambda(1-y)} - (1-y)(1-e^{-\lambda}) \; .$$

Then since $K(y,t)$ is a nonnegative kernel, (2.7) becomes

$$||\mathbf{K}|| \;=\; \sup_{0\le y\le 1} \mathbf{K}1(y) \;=\; 1 - C(\lambda)[1-\ln C(\lambda)] \;, \tag{2.8}$$

where $C(\lambda) = (1-e^{-\lambda})/\lambda$. For small λ, $||\mathbf{K}|| = \lambda^2/8 + O(\lambda^3)$ and the Neumann series does indeed converge rapidly.

If λ is not small, one can still deduce that the Neumann series converges by using (2.8) to prove $||\mathbf{K}|| < 1$. Although $||\mathbf{K}|| \to 1$ as $\lambda \to \infty$, a more refined argument will now show that the convergence is always quite rapid. Because $K(0,t) = 0$, the terms $\mathbf{K}^n(1-y)$ of (2.4) with $n = 1, 2, \ldots$ all vanish at $y = 0$. If c_1, c_2, $c_3, \ldots$ are constants such that $|\mathbf{K}^n(1-y)| \le c_n y$, then

$$|\mathbf{K}^{n+1}(1-y)| \le c_n \mathbf{K}y \le c_n \frac{\mathbf{K}y}{y} y \le \rho c_n y$$

where

$$\rho \;=\; \sup_{0\le y\le 1} \frac{\mathbf{K}y}{y} \;.$$

Then $c_n \le \rho^{n-1} c_1$, so the Neumann series converges as fast as $(\rho + \rho^2 + \cdots)c_1 y$. A numerical calculation of $(\mathbf{K}y)/y$ indicates that $\rho < .3884$ for all λ, with the maximum ρ achieved at $\lambda = 3.39$.

Sections 3 and 4 will give explicit formulas that approximate $Q(y)$ for small and large λ. $Q(y)$ could be calculated for other values of λ from the Neumann series. Although it converges rapidly, each term of the series requires a numerical integration for each value of y. To avoid that problem, Section 5 transforms the integral equation (2.5) to an infinite system of linear algebraic equations and examines the effect of truncating the system to a finite size.

3 Light Traffic Asymptotics

When λ is small most stages serve only one customer; then $P(y)$ is near y. This section finds closer estimates of $P(y)$ and $Q(y)$ in this limiting situation.

Bounds on $Q(y)$ follow easily from (2.5) because $K(y,t)$ is nonnegative. Starting with $Q(y) \geq 0$, (2.5) shows $Q(y) \geq 1-y$, then $Q(y) \geq (\mathbf{I}+\mathbf{K})(1-y)$, etc.; each partial sum of the Neumann series is a lower bound on $Q(y)$. The same argument, starting with $Q(y) \leq 1$, shows $Q(y) \leq 1 - y + \mathbf{K}1(y), \ldots$, and ultimately

$$Q(y) \leq (\mathbf{I}+\mathbf{K}+\cdots+\mathbf{K}^n)(1-y) + \mathbf{K}^{n+1}1(y) \ . \tag{3.1}$$

Truncating the Neumann series at the n^{th} term underestimates $Q(y)$ by at most $\mathbf{K}^{n+1}1(y)$. Simple instances of these bounds are $(\mathbf{I}+\mathbf{K})(1-y) \leq Q(y) \leq (1-y) + \mathbf{K}1(y)$, or

$$e^{-\lambda(1-y)} - e^{-\lambda}(1-y) \leq P(y) \leq \frac{1-e^{-\lambda(1-y)}}{\lambda(1-y)} - \frac{1-e^{-\lambda}}{\lambda}(1-y) \ . \tag{3.2}$$

When λ is small, Section 2 has already shown that $\sup_{0\leq y\leq 1} \mathbf{K}1(y) = ||\mathbf{K}|| = O(\lambda^2)$. The error term in (3.1) is at most $\sup_{0\leq y \leq 1} \mathbf{K}^{n+1}1(y) = O(\lambda^{2n+2})$. To make the bound more precise, write

$$K(y,t) \ = \ \lambda(1-y)\int_{\lambda(1-y)t}^{\lambda t} e^{-x}\,dx \leq \lambda^2 y(1-y)t \ .$$

Then $\mathbf{K}1(y) \leq \lambda^2 y(1-y)/2$. An induction argument, based on $\mathbf{K}^{n+1}1(y) = \mathbf{K}\mathbf{K}^n 1(y)$, proves

$$\mathbf{K}^{n+1}1(y) \leq 6y(1-y)\left(\frac{\lambda^2}{12}\right)^{n+1} \leq \frac{3}{2}\left(\frac{\lambda^2}{12}\right)^{n+1} \ . \tag{3.3}$$

In (3.2) the upper bound on $P(y)$ is accurate to $\sup_{0\leq y\leq 1} \mathbf{K}^2 1(y) \leq \lambda^4/96$. Expanding (3.2) gives

$$Q(t) = 1 - t + \frac{\lambda^2}{6}t(1-t) - \frac{\lambda^3}{24}t(1-t)(2-t) + O(\lambda^4) \ . \tag{3.4}$$

Substitution into (2.6) provides p_k within $O(\lambda^{k+2})$ as $\lambda \to 0$. The leading terms are

$$\begin{cases} p_1 \ = \ 1 - \frac{\lambda^2}{6} + \frac{\lambda^3}{12} + O(\lambda^4) \\ p_k \ = \ \frac{\lambda^k}{(k+1)!} - \frac{(k+1)\lambda^{k+1}}{(k+2)!} + O(\lambda^{k+2}), \quad k \geq 2 \ . \end{cases} \tag{3.5}$$

From (3.5), the mean μ and variance σ^2 of the number served in a stage are

$$\mu = 1 + \frac{\lambda^2}{6} - \frac{\lambda^3}{24} + O(\lambda^4), \tag{3.6}$$

$$\sigma^2 = \frac{\lambda^2}{6} + \frac{\lambda^3}{24} + O(\lambda^4) \ . \tag{3.7}$$

4 Heavy Traffic Asymptotics

With λ large, the lower bound in (3.2) should be accurate because it is close to the generating function for a Poisson distribution with mean λ. Indeed, the extra term $-e^{-\lambda}(1-y)$ is a correction that accounts for the fact that every stage serves at least one customer. This section develops a heavy-traffic approximation for $P(y)$ which leads to asymptotic series for the moments.

It will be convenient to introduce a new variable $\xi = \lambda(1-y)$ and a function $\vartheta(\xi)$ defined by $P(y) = e^{-\xi}\vartheta(\xi)$. Like $P(y)$ and $Q(y) = 1 - P(y)$, $\vartheta(\xi)$ depends on λ as well as ξ. For large λ one expects $P(y)$ to be close to $e^{-\lambda(1-y)}$ and hence $\vartheta(\xi)$ to be near 1. A sequence of approximations of the form

$$P^{(r)}(y) = e^{-\xi}\vartheta^{(r)}(\xi), \quad r \geq 0, \tag{4.1}$$

will now be given, where

$$\vartheta^{(r)}(\xi) = \sum_{k=0}^{r} \frac{\vartheta_k(\xi)}{\lambda^k}, \tag{4.2}$$

with coefficients $\vartheta_k(\xi)$ independent of λ and defined as follows. Substitute $Q(y) = 1-P(y) = 1 - e^{-\xi}\vartheta(\xi)$ into (2.5), change the variable of integration to $\eta = \lambda(1-t)$ and obtain

$$\vartheta(\xi) = \frac{\xi}{\lambda}\int_0^{\lambda}[e^{-(1-\xi/\lambda)\eta} - e^{\xi-\lambda}]\vartheta(\eta)d\eta + 1 - \frac{\xi e^{\xi-\lambda}}{\lambda} \ . \tag{4.3}$$

Now suppose $\sum_{k=0}^{r}\vartheta_k(\xi)/\lambda^k$ were substituted for $\vartheta(\xi)$ in (4.3) and the $\vartheta_k(\xi)$ determined formally by matching coefficients of like powers of $1/\lambda$. In this matching, the functions

$e^{\xi-\lambda}$, that appear twice in (4.3), would contribute only higher order terms. Also, the upper limit of integration could be increased from λ to ∞ without affecting terms $O(\lambda^{-k})$. Now simplify (4.3) according to these observations and replace $\vartheta(\xi)$ by $\vartheta^{(r)}(\xi)$; this defines the desired heavy traffic approximation by

$$\begin{aligned} \vartheta^{(r)}(\xi) &= \frac{\xi}{\lambda}\int_0^\infty e^{-(1-\xi/\lambda)\eta}\vartheta^{(r)}(\eta)d\eta + 1 \\ &= \sum_{j=0}^{\infty}\frac{\xi^{j+1}}{\lambda^{j+1}j!}\int_0^\infty e^{-\eta}\eta^j\vartheta^{(r)}(\eta)d\eta + 1 \ . \end{aligned} \tag{4.4}$$

Matching coefficients in (4.4) leads to $\vartheta_0(\xi) = 1$ and the recurrence

$$\vartheta_{k+1}(\xi) = \sum_{j=0}^{k}\frac{\xi^{j+1}}{j!}\int_0^\infty e^{-\eta}\eta^j\vartheta_{k-j}(\eta)d\eta, \quad k \geq 0 \ . \tag{4.5}$$

The first few $\vartheta_k(\xi)$ from (4.5) are $\vartheta_0(\xi) = 1$, $\vartheta_1(\xi) = \xi$, $\vartheta_2(\xi) = \xi + \xi^2$, $\vartheta_3(\xi) = 3\xi + 2\xi^2 + \xi^3$, $\vartheta_4(\xi) = 13\xi + 8\xi^2 + 3\xi^3 + \xi^4, \ldots$. It is clear from (4.5) that $\vartheta_k(\xi)$ is a polynomial

$$\vartheta_k(\xi) = \sum_{j=0}^{k} c(k,j)\xi^j \ , \tag{4.6}$$

with coefficients $c(k,j)$ satisfying $c(0,0) = 1$, $c(k,0) = 0$, $k \geq 1$, and

$$c(k+1, j+1) = \sum_{m=0}^{k-j} c(k-j, m)\frac{(j+m)!}{j!} \ . \tag{4.7}$$

This series for $\vartheta^{(r)}(\xi)$ yields a corresponding approximation $\mu^{(r)}$ for the mean $\mu = P'(1) = \lambda \lim_{\xi\to 0}\frac{1-e^{-\xi}\vartheta(\xi)}{\xi} = \lambda[1-\vartheta'(0)]$. Replacing $\vartheta(\xi)$ by $\vartheta^{(r)}(\xi)$ gives

$$\mu^{(r)} = \lambda\left[1 - \sum_{k=0}^{r}\frac{\vartheta'_{k+1}(0)}{\lambda^{k+1}}\right] = \lambda\left[1 - \sum_{k=0}^{r}\frac{c(k+1,1)}{\lambda^{k+1}}\right] \ . \tag{4.8}$$

For example, $\mu^{(0)}, \ldots, \mu^{(4)}$ are given by the partial sums in

$$\mu^{(5)} \;=\; \lambda - 1 - \frac{1}{\lambda} - \frac{3}{\lambda^2} - \frac{13}{\lambda^3} - \frac{71}{\lambda^4} \; . \tag{4.9}$$

A similar argument gives approximations for higher moments. For example, standard calculations yield for the variance approximations $v^{(r)}$, $r \le 3$, the partial sums in

$$v^{(3)} \;=\; \lambda + \frac{1}{\lambda} + \frac{6}{\lambda^2} + \frac{39}{\lambda^3} \; . \tag{4.10}$$

The remainder of this section shows that, as $r \to \infty$, (4.8) in fact becomes an asymptotic series for μ. The same techniques can be used to prove similar results for higher moments.

Theorem 4.1 *For all $r \ge 0$, as $\lambda \to \infty$,*

$$|\mu - \mu^{(r)}| \;=\; O\left(\frac{1}{\lambda^{r+1}}\right) \; . \tag{4.11}$$

Proof. Differentiate (2.5) and obtain at $y = 1$

$$\begin{aligned} \mu \;&=\; 1 + \lambda \int_0^1 [1 - e^{-\lambda z}] Q(z) dz \\ &=\; 1 + \lambda \int_0^1 [1 - e^{-\lambda z}] Q^{(r)}(z) dz + \lambda \int_0^1 [1 - e^{-\lambda z}][Q(z) - Q^{(r)}(z)] dz \; . \end{aligned} \tag{4.12}$$

Substituting $\eta = \lambda(1 - z)$, $Q^{(r)}(z) = 1 - e^{-\eta} \sum_{k=0}^{r} \vartheta_k(\eta)/\lambda^k$ and integrating gives for the first two terms

$$\begin{aligned} &1 + \lambda \int_0^1 [1 - e^{-\lambda z}] Q^{(r)}(z) dz = \\ &\qquad \lambda - \sum_{k=0}^{r} \int_0^{\lambda} e^{-\eta} \frac{\vartheta_k(\eta)}{\lambda^k} d\eta + e^{-\lambda} \left[1 + \sum_{k=0}^{r} \int_0^{\lambda} \frac{\vartheta_k(\eta)}{\lambda^k} d\eta \right] \; . \end{aligned} \tag{4.13}$$

Now extend the upper limit of the first integral from λ to ∞; this adds only a term $O(e^{-\lambda})$ to (4.13). The last term in (4.13) is $O(\lambda e^{-\lambda})$, so

$$1 + \lambda \int_0^1 [1 - e^{-\lambda z}] Q^{(r)}(z) dz \;=\; \lambda - \sum_{k=0}^{r} \int_0^{\lambda} e^{-\eta} \frac{\vartheta_k(\eta)}{\lambda^k} d\eta + O(\lambda e^{-\lambda}) \; .$$

Differentiation of (4.5) then shows that, by (4.8),

$$\begin{aligned} 1 + \lambda \int_0^1 [1 - e^{-\lambda z}] Q^{(r)}(z) dz \;&=\; \lambda \left[1 - \sum_{k=0}^{r} \frac{\vartheta'_{k+1}(0)}{\lambda^{k+1}} \right] + O(\lambda e^{-\lambda}) \\ &=\; \mu^{(r)} + O(\lambda e^{-\lambda}) \; . \end{aligned}$$

Substituting into (4.12) gives

$$\begin{aligned} \mu - \mu^{(r)} \;&=\; \lambda \int_0^1 [1 - e^{-\lambda z}][Q(z) - Q^{(r)}(z)] dz + O(\lambda e^{-\lambda}) \\ &\leq\; \lambda \sup_{0 \leq z \leq 1} |Q(z) - Q^{(r)}(z)| + O(\lambda e^{-\lambda}) \; . \end{aligned} \tag{4.14}$$

It is shown below that

$$\lambda \sup_{0 \leq z \leq 1} |Q(z) - Q^{(r)}(z)| \;=\; O\left(\frac{1}{\lambda^{r-1}} \right) \; . \tag{4.15}$$

Then, since $\mu^{(r)} = \mu^{(r+2)} + O(1/\lambda^{r+1})$, the theorem follows at once from (4.14) and (4.15) with r replaced by $r + 2$.

To prove (4.15), first introduce the function $\Gamma_r(y)$ defined by

$$Q^{(r)}(y) \;=\; (\mathbf{K} Q^{(r)})(y) + (1 - y) + \Gamma_r(y) \; . \tag{4.16}$$

Then

$$Q^{(r)}(y) - Q(y) \;=\; (\mathbf{K}\{Q^{(r)} - Q\})(y) + \Gamma_r(y)$$

and

$$\sup_{0\le y\le 1} |Q^{(r)}(y) - Q(y)| \le \frac{1}{1-||\mathbf{K}||} \sup_{0\le y\le 1} \Gamma_r(y) .$$

By (2.8)

$$1 - ||K|| < C(\lambda)[1-\ln C(\lambda)] \;=\; \frac{\ln\lambda}{\lambda} + O(e^{-\lambda}) .$$

Then (4.15) and hence the theorem will follow if

$$\sup_{0\le y\le 1} \Gamma_r(y) \;=\; O\left(\frac{1}{\lambda^{r+1}}\right) . \tag{4.17}$$

The proof concludes by showing (4.17) directly.

Substituting for $Q^{(r)}$ in (4.16) gives, after routine manipulations,

$$\Gamma_r(y) = e^{-\xi}\left[1 - \sum_{k=0}^{r}\frac{\vartheta_k(\xi)}{\lambda^k}\right] + \frac{\xi}{\lambda}e^{-\xi}\sum_{k=0}^{r}\int_0^\lambda e^{\eta\xi/\lambda}e^{-\eta}\frac{\vartheta_k(\eta)}{\lambda^k}d\eta$$
$$-\frac{\xi}{\lambda}e^{-\lambda}\left[1 + \sum_{k=0}^{r}\int_0^\lambda \frac{\vartheta_k(\eta)}{\lambda^k}d\eta\right] . \tag{4.18}$$

Recall that the $\vartheta_k(\xi)$ are polynomials of degree k. The last term in (4.18) is $O(\lambda e^{-\lambda})$ for $0 \le \xi \le \lambda$, so expansion of $e^{\xi\eta/\lambda}$ and rearrangement gives

$$\Gamma_r(y) = e^{-\xi}\frac{\xi}{\lambda}\sum_{k=0}^{r}\sum_{j=0}^{\infty}\int_0^\lambda e^{-\eta}\frac{(\xi\eta/\lambda)^j}{j!}\frac{\vartheta_k(\eta)}{\lambda^k}d\eta - e^{-\xi}\sum_{k=1}^{r}\frac{\vartheta_k(\xi)}{\lambda^k} + O(\lambda e^{-\lambda}) . \tag{4.19}$$

To estimate the terms in (4.19) note that the recurrence for ϑ_k in (4.5) gives

$$\sum_{k=1}^{r}\frac{\vartheta_k(\xi)}{\lambda^k} \;=\; \sum_{k=0}^{r-1}\frac{\vartheta_{k+1}(\xi)}{\lambda^{k+1}} \;=\; \frac{\xi}{\lambda}\sum_{k=0}^{r-1}\sum_{j=0}^{k}\int_0^\infty e^{-\eta}\frac{(\xi\eta/\lambda)^j}{j!}\frac{\vartheta_{k-j}(\eta)}{\lambda^{k-j}}d\eta,$$

whereupon reorganizing the double sum yields

$$\sum_{k=1}^{r}\frac{\vartheta_k(\xi)}{\lambda^k} \;=\; \frac{\xi}{\lambda}\sum_{k=0}^{r-1}\sum_{j=0}^{r-k-1}\int_0^\infty e^{-\eta}\frac{(\xi\eta/\lambda)^j}{j!}\frac{\vartheta_k(\eta)}{\lambda^k}d\eta . \tag{4.20}$$

Now consider the terms of the double sum in (4.19) for which $0 \le k \le r-1$, and $0 \le j < r-k$. If the upper limit of the integral is increased from λ to ∞ for these terms, then by (4.20) these terms are all cancelled by the single sum in (4.19). It is easily verified that the above change in the integral creates a change of at most $O(e^{-\lambda})$ in $\Gamma_r(y)$, so the terms remaining after the cancellation give

$$\Gamma_r(y) = e^{-\xi}\frac{\xi}{\lambda}\left\{\int_0^\lambda e^{-\eta(1-\xi/\eta)}\frac{\vartheta_r(\eta)}{\lambda^r}d\eta + \sum_{k=0}^{r-1}\sum_{j=r-k}^{\infty}\int_0^\lambda e^{-\eta}\frac{(\xi\eta/\lambda)^j}{j!}\frac{\vartheta_k(\eta)}{\lambda^k}d\eta\right\} + O(\lambda e^{-\lambda})\,.$$

Introducing the bound $\sum_{j\ge r-k}\frac{(\xi\eta/\lambda)^j}{j!} \le e^{\xi\eta/\lambda}\frac{(\xi\eta/\lambda)^{r-k}}{(r-k)!}$ in the double sum gives

$$\Gamma_r(y) = e^{-\xi}\frac{\xi}{\lambda}\left\{\int_0^\lambda e^{-\eta(1-\xi/\eta)}\frac{\vartheta_r(\eta)}{\lambda^r}d\eta + \sum_{k=0}^{r-1}\int_0^\lambda e^{-\eta(1-\xi/\lambda)}\frac{(\xi\eta)^{r-k}}{(r-k)!}\frac{\vartheta_k(\eta)}{\lambda^r}d\eta\right\} + O(\lambda e^{-\lambda})\,. \tag{4.21}$$

Estimates of this expression are easily obtained from estimates of the function

$$g(\lambda) \;=\; \sup_{0\le\alpha\le 1}\alpha^i e^{-\alpha\lambda}\int_0^\lambda e^{-\eta(1-\alpha)}\eta^j d\eta \tag{4.22}$$

where $i \ge 1$ and $j \ge 0$ are integers. It is easily shown that $g(\lambda) = O(\lambda^{-i})$ (the hidden multiplicative constant depends on i and j). Use of (4.22) in (4.21) with $\alpha = \xi/\lambda$ easily shows that both terms in (4.21) are no greater than the desired error term $O(1/\lambda^{r+1})$; the routine details are left to the reader. ■

5 Intermediate λ

At values of λ too large for Section 3 and too small for Section 4, a numerical procedure is needed. The integral equation (2.5) can be transformed into a system of simultaneous linear algebraic equations. A straightforward way would be to substitute $Q(y) = 1-P(y) = 1-p_1y-p_2y^2\ldots$ into (2.5) and match coefficients of like powers of y. The resulting system, to be solved for the unknowns p_1, $p_2,\ldots$, would not have a symmetric matrix. That would complicate the problem of estimating the error introduced by truncating the system to one with only finitely many unknowns $p_1, \ldots, p_N$. The procedure that follows has the advantage of a symmetric matrix.

Begin by expanding the kernel (2.2) in the form

$$K(y,t) \;=\; \lambda(1-y)e^{-\lambda t}\sum_{k=1}^{\infty}\frac{\lambda^i y^i t^i}{i!}\;,$$

so that (2.5) is

$$Q(y) \;=\; \lambda(1-y)\sum_{i=1}^{\infty}\frac{\lambda^i y^i}{i!}\int_0^1 e^{-\lambda t}t^i Q(t)dt + 1 - y\;. \tag{5.1}$$

Multiplying both sides by $e^{-\lambda t}t^k$ and integrating will convert (5.1) to an algebraic system for unknowns that are integrals of $e^{-\lambda y}y^k Q(y)$. If the unknowns are properly normalized the system can be made symmetric. Accordingly, define unknowns

$$a_k \;=\; \left(\frac{\lambda^k}{k!}\right)^{1/2}\int_0^1 e^{-\lambda y}y^k Q(y)dy \tag{5.2}$$

and the vector $\mathbf{a} = (a_1, a_2, \ldots)$. The system for $\mathbf{a}$ is

$$\mathbf{a} \;=\; \mathbf{T}\mathbf{a} + \mathbf{b} \tag{5.3}$$

where $\mathbf{b} = (b_1, b_2, \ldots)$,

$$b_k \;=\; \left(\frac{\lambda^k}{k!}\right)^{1/2}\int_0^1 e^{-\lambda y}y^k(1-y)dy\;, \tag{5.4}$$

and $\mathbf{T}$ is the infinite symmetric matrix with elements

$$T_{kj} \;=\; \lambda\left(\frac{\lambda^{k+j}}{k!j!}\right)^{1/2}\int_0^1 e^{-\lambda t}t^{k+j}(1-t)dt \tag{5.5}$$

for $k \geq 1$ and $j \geq 1$. When (5.3) is solved for $\mathbf{a}$, (5.1) will be

$$Q(y) \;=\; \lambda(1-y)\sum_{i=1}^{\infty}\left(\frac{\lambda^i}{i!}\right)^{1/2}a_i y^i + 1 - y\;.$$

As in (2.6), matching coefficients gives

$$p_k = \begin{cases} 1 - \lambda^{3/2} a_1, & k = 1, \\ \left(\frac{\lambda^{k+1}}{(k-1)!}\right)^{1/2} a_{k-1} - \left(\frac{\lambda^{k+2}}{k!}\right)^{1/2} a_k, & k \geq 2. \end{cases}$$

Then a simpler formula relates a_k to the tail probability

$$\sum_{i=k+1}^{\infty} p_i = \lambda \left(\frac{\lambda^k}{k!}\right)^{1/2} a_k, \qquad k \geq 1 . \tag{5.6}$$

An iterative procedure for solving (5.3) leads to

$$\mathbf{a} = \sum_{r=0}^{\infty} \mathbf{T}^r \mathbf{b} . \tag{5.7}$$

A proof that (5.7) actually converges will follow by showing that $\mathbf{T}$ has spectral radius $\rho(\mathbf{T}) < 1$.

Theorem 5.1 $\mathbf{T}$ *is a positive-definite, self-adjoint, Hilbert-Schmidt (hence compact) linear operator in* l^2 *with spectral radius* $\rho(\mathbf{T}) < .505$

Proof. $\mathbf{T}$ is self-adjoint because it is real and symmetric. To prove $\mathbf{T}$ is Hilbert-Schmidt, note that, from (5.5),

$$0 < T_{kj} < \lambda \left(\frac{\lambda^{k+j}}{k!j!}\right)^{1/2} .$$

Then

$$trace(\mathbf{T}^2) = \sum_{k,j} T_{kj}^2 < \lambda^2 \sum \frac{\lambda^k}{k!} \frac{\lambda^j}{j!} = \lambda^2 e^{2\lambda} < \infty .$$

Positive-definiteness will follow by writing (5.5) as

$$T_{kj} = \int_0^1 f_k(t) f_j(t) dt$$

with

$$f_k(t) \;=\; \left\{ \frac{\lambda^{k+1}(1-t)}{k!} \right\}^{1/2} t^k e^{-\lambda t/2} \;.$$

For any $\mathbf{x}$ in l^2,

$$\langle \mathbf{x},\ \mathbf{T}\mathbf{x} \rangle \;=\; \sum_{k,j} \bar{x}_k x_j \int_0^1 f_k(t) f_j(t) dt \;=\; \int_0^1 \left| \sum x_j f_j(t) \right|^2 dt \geq 0 \tag{5.8}$$

and $\langle \mathbf{x},\ \mathbf{T}\mathbf{x} \rangle = 0$ only if $\mathbf{x} = \mathbf{0}$.

Since $\mathbf{T}$ is positive-definite and self-adjoint, $\rho(\mathbf{T}) = \sup_{||\mathbf{x}||=1} \langle \mathbf{x},\ \mathbf{T}\mathbf{x} \rangle$. Applying the Cauchy-Schwarz inequality to the integrand (5.8) gives

$$\rho(\mathbf{T}) \;\leq\; \int_0^1 \sum_{k=1}^{\infty} |f_k(t)|^2 dt \;=\; \lambda \int_0^1 (1-t) e^{-\lambda t} (e^{\lambda t^2} - 1) dt \;,$$

$$\rho(\mathbf{T}) \;\leq\; \lambda e^{-\lambda/4} \int_0^{1/2} e^{\lambda u^2} du - (e^{-\lambda} - 1 + \lambda)/\lambda \;, \tag{5.9}$$

the last line being obtained by the substitution $t = u + \frac{1}{2}$. It is easy to verify that $\rho(\mathbf{T}) \to 0$ both as $\lambda \to 0$ and as $\lambda \to \infty$. An analytical proof of $\rho(\mathbf{T}) < 1$ for all $\lambda \geq 0$ is also not difficult. However, the following simple numerical approach yields a good result.

For any c in $0 < c < 1/2$ let the range of integration in (5.9) be broken into two parts, $[0, c]$ and $[c, 1/2]$. In each part, bound u^2 by a linear function, $u^2 \leq cu$ in $[0, c]$ and $u^2 \leq \left(\frac{1}{2} + c\right) u - \frac{c}{2}$ in $[c, 1/2]$. Then (5.9) simplifies to

$$\rho(\mathbf{T}) \leq \frac{1-2c}{1+2c} + \frac{e^{-\lambda\left(\frac{1}{4} - c^2\right)}}{c(2c+1)} - \frac{e^{-\lambda/4}}{c} + \frac{1 - e^{-\lambda}}{\lambda} \;. \tag{5.10}$$

The best bound is obtained from (5.10) with c chosen as a function of λ, but $c = .3$ gives $\rho(\mathbf{T}) < .505$ for every λ. ∎

Numerical integration in (5.9) showed that $\rho(\mathbf{T}) < .41$.

Although the series solution (5.7) converges to $\mathbf{a}$ as rapidly as the power series $\sum \rho(\mathbf{T})^k$, it contains products of infinite matrices. To obtain a finite computation, (5.3) will be approximated by an $N \times N$ system. As a preliminary step, approximate $\mathbf{T}$ by a matrix $\mathbf{T}^{(N)}$, having the same elements T_{kj} in the principal minor with $1 \le k \le N$ and $1 \le j \le N$, but having zeros elsewhere. Then approximate (5.3) by the system

$$\hat{\mathbf{a}} \; = \; \mathbf{T}^{(N)}\hat{\mathbf{a}} + \mathbf{b} \tag{5.11}$$

for the unknown vector $\hat{\mathbf{a}}$. Because rows $N+1, N+2, \ldots$ of $\mathbf{T}^{(N)}$ contain only zeros, $\hat{\mathbf{a}}$ will have the form $(\hat{a}_1, \ldots, \hat{a}_N, b_{N+1}, b_{N+2}, \ldots)$. The first N equations of the system (5.11) are then an $N \times N$ system for $\hat{a}_1, \ldots, \hat{a}_N$.

To study the truncation error, combine (5.3) and (5.11) into a system

$$\mathbf{a} - \hat{\mathbf{a}} \; = \; \mathbf{T}(\mathbf{a} - \hat{\mathbf{a}}) + (\mathbf{T} - \mathbf{T}^{(N)})\hat{\mathbf{a}} \tag{5.12}$$

for the error vector $\mathbf{a} - \hat{\mathbf{a}}$. The ratio $|\mathbf{a} - \hat{\mathbf{a}}|/|\hat{\mathbf{a}}|$ of l^2 norms of $\mathbf{a} - \hat{\mathbf{a}}$ and $\hat{\mathbf{a}}$ will be a measure of the relative error caused by truncation.

Theorem 5.2 *The relative truncation error satisfies*

$$\{1 - \rho(\mathbf{T})\}|\mathbf{a} - \hat{\mathbf{a}}|/|\hat{\mathbf{a}}| \le \sqrt{trace\{(\mathbf{T} - \mathbf{T}^{(N)})^2\}} \le \sqrt{2 \sum_{j \ge N+1} \sum_{k \ge 1} T_{kj}^2} \; . \tag{5.13}$$

Proof. Apply the triangle inequality to (5.12) and obtain $|\mathbf{a} - \hat{\mathbf{a}}| \le |\mathbf{T}(\mathbf{a} - \hat{\mathbf{a}})| + |(\mathbf{T} - \mathbf{T}^{(N)})\hat{\mathbf{a}}|$. The first inequality of (5.13) follows because the operator $\mathbf{T} - \mathbf{T}^{(N)}$ has norm less than $\sqrt{trace\{(\mathbf{T} - \mathbf{T}^{(N)})^2\}}$. The trace is the sum of squares of elements of $\mathbf{T}$ outside the principal $N \times N$ minor. The second inequality of (5.13) bounds this sum in a way that includes some elements twice. ■

The integral formula (5.5) for T_{kj} is not convenient for calculations. Introduce

$$J_n = \frac{\lambda^{n+1}}{n!} \int_0^1 e^{-\lambda t} t^n \, dt \; .$$

Since $J_0 = 1 - e^{-\lambda}$ and $J_n = J_{n-1} - e^{-\lambda}\lambda^n/n!$,

$$J_n = 1 - e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=n+1}^{\infty} \frac{\lambda^k}{k!} \; . \tag{5.14}$$

Express the integral in (5.5) in terms of J_n and J_{n+1} to obtain

$$T_{kj} = e^{-\lambda} \sqrt{\frac{\lambda^{k+j-2}}{k!j!}} \sum_{r \geq k+j+2} \frac{(r-k-j-1)\lambda^{r-k-j}}{r(r-1)\cdots(k+j+1)} \; . \tag{5.15}$$

The finite sum (5.14) could have been used to express T_{kj} in finite terms, but the series (5.15) converges rapidly and is less sensitive to roundoff errors.

Table 1 gives bounds, obtained from (5.13), on $|\mathbf{a} - \hat{\mathbf{a}}|/|\hat{\mathbf{a}}|$. The calculations used $\rho(\mathbf{T}) <$ .41 for all λ. The table shows that good accuracy is obtainable from only moderately large truncated systems.

Starting from (5.13), the following explicit, though crude bound is proved in the Appendix:

$$\frac{|\mathbf{a} - \hat{\mathbf{a}}|}{|\hat{\mathbf{a}}|} \leq \frac{1}{1 - \rho(\mathbf{T})} \left(\frac{2}{\pi N} \right)^{1/4} \frac{8\lambda - 32 + (32 - \lambda^2)e^{-\lambda/4}}{\lambda^{3/2}} \; . \tag{5.16}$$

This bound shows that, for all $N \geq 1$, the relative error tends to 0 both as $\lambda \to 0$ and as $\lambda \to \infty$. It also verifies that the relative error tends to 0 uniformly in λ as $N \to \infty$.

Table 1: Bound on relative error $|\mathbf{a} - \hat{\mathbf{a}}|/|\hat{\mathbf{a}}|$ caused by truncating (5.3) to an $N \times N$ system.

$N \setminus \lambda$	2	4	8	14	20
2	.0958	.3041	.4648	.3133	.1835
4	.0207	.1188	.2982	.2614	.1662
6	.0037	.0404	.1799	.2243	.1600
8	.0006	.0120	.0988	.1857	.1539
10	.0001	.0031	.0490	.1448	.1449
14		.0002	.0090	.0707	.1149
18			.0011	.0256	.0743
22			.0001	.0070	.0381
30				.0002	.0051
40					.0001

References

[1] B. Avi-Itzhak and S. Halfin, "Response Times in Gated M/G/1 Queues: The Processor-Sharing Case," *Queueing Systems*, **4** (1989), 263-279.

[2] S. Browne, E. G. Coffman, Jr. E. N. Gilbert, and P. E. Wright, "Gated, Exhaustive, Parallel Service," preprint (1990), AT&T Bell Laboratories, Murray Hill, NJ 07974.

[3] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. I*, Interscience, New York, 1953.

[4] D. L. Jagerman and B. Sengupta, "A Functional Equation Arising in a Queue with a Gating Mechanism," *Probability in the Engineering and Information Sciences*, **3** (1989), 417-433.

[5] C. Knessl and J. B. Keller, "Asymptotic Properties of Eigenvalues of Integral Equations," *SIAM Journal on Applied Mathematics*, **51** (1991), 214-232.

[6] W. E. Olmstead and A. K. Gautesen, "Asymptotic Solution of Some Singularly Perturbed Fredholm Integral Equations," *Journal of Applied Mathematics and Physics*, **40** (1989), 230-244.

[7] K. M. Rege and B. Sengupta, "A Single-Server Queue with Gated Processor-Sharing Discipline," *Queueing Systems*, **4** (1989), 249-261.

[8] D. Slepian, "A Numerical Method for Determining the Eigenvalues and Eigenfunctions of Analytic Kernels," *SIAM Journal on Numerical Analysis*, **5** (1968), 586-600.

[9] A. E. Taylor and D. C. Lay, *Introduction to Functional Analysis*, John Wiley & Sons, New York, 1980.

Appendix

Proof of (5.16)

The trace in (5.13) is a sum of squares of the elements of $\mathbf{T} - \mathbf{T}^{(N)}$, i.e., a sum $\sum T_{kj}^2$ where only elements outside the principal $N \times N$ minor are summed. From (5.5),

$$\begin{aligned} trace\{(\mathbf{T} - \mathbf{T}^{(N)})^2\} &< 2\sum_{j=N+1}^{\infty} \sum_{k=1}^{\infty} T_{kj}^2 \\ &= 2\lambda^2 \int_0^1 \int_0^1 e^{-\lambda(y+z)}(1-y)(1-z) \sum_{j=N+1}^{\infty} \frac{(\lambda yz)^j}{j!} \sum_{k=1}^{\infty} \frac{(\lambda yz)^t}{k!} dydz \ . \end{aligned} \tag{A1}$$

In (A1) the sum on k is $e^{\lambda yz} - 1 < e^{\lambda z}$. The sum on j is

$$\sum_{j=N+1}^{\infty} \frac{(\lambda yz)^j}{j!} = \lambda yz e^{\lambda yz} \int_0^1 e^{-\lambda xyz} \frac{(\lambda xyz)^N}{N!} dx \ , \tag{A2}$$

as repeated integration by parts will verify. The integrand of (A2) is greatest at $x = N/(\lambda yz)$, so the integral in (A2) is bounded by $(N/e)^N/N! < 1/\sqrt{2\pi N}$. These bounds simplify (A1) to

$$trace\{(\mathbf{T} - \mathbf{T}^{(N)})^2\} < \lambda^3 \sqrt{\frac{2}{\pi N}} \int_0^1 \int_0^1 e^{-\lambda(y+z-2yz)}(1-y)(1-z)yzdydz \ .$$

In the exponent, $2yz < y^2 + z^2$, so

$$trace\{(\mathbf{T} - \mathbf{T}^{(N)})^2\} < \lambda^3 \sqrt{\frac{2}{\pi N}} \left| 2\int_0^{1/2} e^{-\lambda y(1-y)} y(1-y)dy \right|^2 \ .$$

Replace $y(1-y)$ in the exponent by a linear lower bound $y/2$; then evaluate the integral to complete the proof of (5.16). ■