Decompositions of the $M/M/1$ transition function

Joseph Abate
900 Hammond Road, Ridgewood, NJ 07450, USA

Masaaki Kijima
Graduate School of Systems Management, The University of Tsukuba, Otsuka, Bunkyo-ku, Tokyo 112, Japan

Ward Whitt
AT&T Bell Laboratories, Room 2C-178, Murray Hill, NJ 07974-0636, USA

Received 13 July 1990; revised 30 January 1991

Two decompositions are established for the probability transition function of the queue length process in the $M/M/1$ queue by a simple probabilistic argument. The transition function is expressed in terms of a zero-avoiding probability and a transition probability to zero in two different ways. As a consequence, the $M/M/1$ transition function can be represented as a positive linear combination of convolutions of the busy-period density. These relations provide insight into the transient behavior and facilitate establishing related results, such as inequalities and asymptotic behavior.

Keywords: $M/M/1$ queue; transient behavior; queue length process; simple random walk; zero-avoiding probabilities; relaxation time.

1. Introduction

Let $Q(t)$ represent the number in system at time t in an $M/M/1$ queue with service rate 1 and arrival rate ρ, $0<\rho<\infty$. Let $P_{ij}(t)$ be the transition probability

$$P_{ij}(t)=P(Q(t)=j \mid Q(0)=i). \tag{1}$$

We wish to express $P_{ij}(t)$ in terms of more elementary building blocks. In particular, as we and many others have done before (see pp. 5–17 of Prabhu [16], Baccelli and Massey [7], p. 163 of Abate and Whitt [2] and p. 341 of Abate and Whitt [3]), we use related unrestricted and absorbing processes. In section 2 we provide background and in section 3 we establish two decompositions, i.e., two representations of $P_{ij}(t)$ in terms of the probability transition functions of the unrestricted and absorbing processes. In the remaining sections we deduce consequences of the decompositions and establish related results. Sections 4 and 5 contain new representations and inequalities, while section 6 describes the asymptotic behavior as $t \to \infty$. In section 6 we include a description of the asymptotic behavior with a general geometric initial distribution. As noted by Massey [15], this is interesting because it shows that the relaxation time parameter actually depends on the initial distribution, even though it does not for the initial distributions with compact support.

2. Background

The *unrestricted process*, say $X(t)$, is a birth-and-death process on all the integers with birth rate ρ and death rate 1. Let its transition probability be

$$U_{j-i}(t) = P(X(t) = j \mid X(0) = i). \tag{2}$$

We are justified in writing $U_{j-i}(t)$ as a function of $j-i$ because $X(t)$ is spatially homogeneous. It is well known that $U_j(t) = \rho^j U_{-j}(t)$ and

$$U_{-j}(t) = \rho^{-j/2} \, e^{-(1+\rho)t} I_j\left(2\sqrt{\rho}\, t\right), \tag{3}$$

where $I_j(t)$ is the modified Bessel function; see p. 377 of Abramowitz and Stegun [6] and p. 9 of Prabhu [16].

The *absorbing process*, say ${}^0Q(t)$, is the queue length process $Q(t)$ modified to be absorbing in state 0. Let its transition probability be

$${}^0P_{ij}(t) = P\left({}^0Q(t) = j \mid {}^0Q(0) = i\right). \tag{4}$$

For $i, j > 0$, ${}^0P_{ij}(t)$ is also called the zero-avoiding transition probability or the busy period transition probability, because

$${}^0P_{ij}(t) = P\left(Q(t) = j, \inf_{0 \leqslant s \leqslant t} \{Q(s)\} > 0 \mid Q(0) = i\right). \tag{5}$$

It is well known that the zero-avoiding probabilities can be expressed in terms of the transition probabilities of the unrestricted process by

$${}^0P_{ij}(t) = U_{j-i}(t) - \rho^{-i} U_{-(j+i)}(t). \tag{6}$$

Formula (6) was first obtained by Bailey [8] by transform inversion. A probabilistic derivation was provided by (1.63) of Prabhu [16]. A related probabilistic derivation is theorem 3 of Baccelli and Massey [7].

To put our new results in perspective, we describe some of our previous results. In theorem 4.3 of [2] we expressed the Laplace transform $\hat{P}_{ij}(s)$ of $P_{ij}(t)$ in terms of the Laplace transform $\hat{P}_{0j}(s)$ of the transition probability from 0,

$P_{0j}(t)$, and the Laplace transform $\hat{f}_{0i}(s)$ of the density $f_{0i}(t)$ of the first passage time from 0 to i, i.e.,

$$\hat{P}_{ij}(s) = \hat{P}_{0j}(s)/\hat{f}_{0i}(s). \tag{7}$$

Moreover, in corollary 4.22 of [2] we showed that the transition probabilities from 0 have a relatively simple expression in terms of the $M/M/1$ busy-period cdf, i.e.,

$$P_{0j}(t) = \rho^j F_{j0}(t) - \rho^{j+1} F_{j+1,0}(t), \tag{8}$$

where $F_{j0}(t)$ is the cdf of the first passage time from j to 0, which is just the j-fold convolution of the busy-period cdf. Of course, if $\rho > 1$, then $F_{j0}(t)$ is a defective cdf. The density $f_{j0}(t)$ is directly related to $U_{-j}(t)$ by

$$f_{j0}(t) = (j/t)U_{-j}(t); \tag{9}$$

see (1.65) of [16]. Unfortunately, however, (7) does not seem so useful because $\hat{f}_{0i}(s)$ is complicated. (We do know that $f_{0i}(s) = 1/q_i(-s)$, where $q_i(x)$ is the ith orthogonal polynomial in the spectral representation; see p. 378 of [13] and theorem 8.1 of [3].) Relation (8) is very useful though, and we obtain a new proof via the decompositions; see remark 6 below.

3. The decompositions

Now we establish decompositions that express $P_{ij}(t)$ directly in terms of a zero-avoiding transition probability and a transition probability to 0. We focus on $P_{n+k,n}(t)$, because

$$P_{n,n+k}(t) = \rho^k P_{n+k,n}(t) \quad \text{and} \quad {}^0P_{n,n+k}(t) = (\rho^k)^0 P_{n+k,n}(t) \tag{10}$$

by reversibility. (For $\rho \geqslant 1$, consider the finite state space model on the integers from 0 to N, so that there is a proper limiting distribution, and then let $N \to \infty$. For the zero-avoiding probabilities, let the birth rate in state 0 approach 0 in the original model, so that $P_{ij}(t)$ approaches ${}^0P_{ij}(t)$.) By (3), (6), (8) and (9), the two components are relatively tractable. It is significant that we establish our main decomposition result by a simple probabilistic argument.

THEOREM 1

For all $t \geqslant 0$ and all integers $k \geqslant 0$ and $n \geqslant 0$,

(a) $P_{n+k,n}(t) = {}^0P_{n+k,n}(t) + \rho^n P_{2n+k,0}(t)$

(b) $P_{n+k,n}(t) = {}^0P_{n+k+1,n+1}(t) + \rho^n P_{2n+k+1,0}(t)$.

Proof

(a) Partition the event whose probability we are computing into two subsets, one in which 0 is never hit and the other in which it is. Further partition the event in which 0 is hit according to the first time that it is hit. Hence, we have

$$P_{n+k,n}(t) = {}^0P_{n+k,n}(t) + f_{n+k,0}(t) * P_{0n}(t), \tag{11}$$

where $*$ denotes convolution. Next note that $P_{0n}(t) = \rho^n P_{n0}(t)$ by (10), $f_{n+k,0}(t) = f_{2n+k,n}(t)$ and, conditioning on the first visit to n,

$$P_{2n+k,0}(t) = f_{2n+k,n}(t) * P_{n0}(t). \tag{12}$$

(b) Now assume that $Q(t)$ is defined in terms of the unrestricted process $X(t)$ by imposing an impenetrable barrier at 0, as on p. 11 of [16]. Then partition the event whose probability we are computing into two subsets, one in which 0 is hit and there is a potential departure before the next arrival, and the other the complement. The time such a potential departure first occurs has the same distribution as the first passage time from $n+k+1$ to 0. The transition from $n+k$ to n with this potential departure not occurring has the probability ${}^0P_{n+k+1,n}(t)$. Hence

$$P_{n+k,n}(t) = {}^0P_{n+k+1,n+1}(t) + f_{n+k+1,0}(t) * P_{0n}(t). \tag{13}$$

The rest follows as for part (a). □

Remarks

(1) Part (a) of theorem 1 seems to be new, but a relation equivalent to part (b) appears in (2.1.10) on p. 17 of Conolly [10] with a different proof. A relation equivalent to part (a) has been derived independently by T. Kissinger (private communication). These equivalent relations are in terms of the unrestricted process instead of the zero-avoiding probabilities, but are connected by (6).

(2) Theorem 1 might lead us to conjecture that

$$P_{n+k,n}(t) = {}^0P_{n+k+j,n+j}(t) + \rho^n P_{2n+j,0}(t)$$

for $j = -1$ or $j = +2$, but these are not valid.

(3) Theorem 1 seems to be useful primarily for providing structural insight. From theorem 1 and (3), (6), (8) and (10), we obtain expressions in terms of the modified Bessel functions such as on pp. 82–83 of Cohen [9], but there are differences. It does not seem easy to relate the results. For numerical calculations, we would use the trigonometric integral representations, as indicated in [4]. It is possible to deduce theorem 1 from the trigonometric integral representations, but we do not have an easy proof. □

4. Consequences: representations and inequalities

Theorem 1 is applied in Abate and Whitt [5] to help explain the shape of $P_{n+k,n}(t)$. In particular, contrary to the conjecture on p. 171 of [2], it is shown that the derivative $P'_{n+k,n}(t)$ can have up to three zeros. Now we deduce several other sequences of theorem 1.

COROLLARY 1

For all $t \geqslant 0$ and all integers $k \geqslant 0$ and $n \geqslant 0$,

(a) $$P_{k0}(t) - P_{k+1,0}(t) = {}^0P_{k+1,1}(t) = f_{k+1,0}(t) = U_{-k}(t) - \rho U_{-(k+2)}(t) = \frac{(k+1)}{t} U_{-(k+1)}(t) > 0.$$

(b) $${}^0P_{n+k+1,n+1}(t) - {}^0P_{n+k,n}(t) = \rho^n[P_{2n+k,0}(t) - P_{2n+k+1,0}(t)] = \rho^n[{}^0P_{2n+k+1,1}(t)] = \rho^n f_{2n+k+1,0}(t) > 0,$$

(c) $$P_{n+k,n}(t) - P_{n+k+1,n+1}(t) = \rho^n P_{2n+k+1,0}(t) - \rho^{n+1} P_{2n+k+2,0}(t),$$

(d) $$P_{n+k,n}(t) - P_{n+k+1,n+1}(t) > 0 \text{ if } \rho \leqslant 1.$$

Proof

For the first relation in part (a), let $n = 0$ in theorem 1(b). For the second relation in (a), use first principles: To go from $k+1$ to 0 for the first time, you must go to 1 without hitting 0 and then make a transition to 0, as noted on p. 344 of [3]. For the third relation in (a), apply (6). For the final relation in (a), use (3) and the Bessel function relation $I_k(z) - I_{k+2}(z) = (2(k+1)/z)I_{k+1}(z)$; see 9.6.26 of Abramowitz and Stegun [6]. For a general probabilistic treatment, see p. 81 of Prabhu [17]. For the first relation in part (b), subtract (a) from (b) in theorem 1. For the rest of (b), apply part (a) here. For (c), subtract (a) from (b) in theorem 1 using $n+1$ for n in (a). For (d), apply (c). □

Remarks

(4) The relationship between $P_{k0}(t) - P_{k+1,0}(t)$ and $U_{-k}(t) - \rho U_{-(k+2)}(t)$ is (2.1.8) of Conolly [10]. It also follows from (1.62) and (1.65) of Prabhu [16]. Part (d) can be established by a simple coupling argument after applying (9), as in lemma 10.1 of [2]. Only $P_{n,n+k}(t) \leqslant P_{0k}(t)$ was stated there.

We obtain further relations by summing the equations in corollary 1.

COROLLARY 2

For all $t \geqslant 0$ and all integers $k \geqslant 0$ and $n \geqslant 0$,

(a) $$P_{k0}(t) = \sum_{j=k+1}^{\infty} {}^0P_{j1}(t) = \sum_{j=k+1}^{\infty} f_{j0}(t) = \sum_{j=k+1}^{\infty} [U_{-j}(t) - \rho U_{-(j+2)}(t)],$$

(b) $${}^0P_{n+k,n}(t) = \sum_{j=0}^{n-1} \rho^j[{}^0P_{2j+k+1,1}(t)] = \sum_{j=0}^{n-1} \rho^j f_{2j+k+1,0}(t).$$

Remark

(5) The first two relations in (a) also appear in (1.62) of [16].

We now show that the transition probability $P_{n+k,n}(t)$ can be represented as a positive linear combination of convolutions of the busy-period density.

COROLLARY 3

For all $t \geqslant 0$ and all integers $k \geqslant 0$ and $n \geqslant 0$,

$$P_{n+k,n}(t) = \left[\sum_{j=0}^{n-1} \rho^j f_{2j+k-1,0}(t) + \rho^n \sum_{j=2n+k+1}^{\infty} f_{j0}(t) \right].$$

Proof

Combine theorem 1 (a) and corollary 2. □

Remark

(6) We can express $P_{n+k,n}(t)$ solely in terms of the unrestricted transition probabilities by combining corollary 3 with (9).

Next we obtain a relation between the first-passage-time densities and their cdf's. We give two proofs, one invoking corollary 2 and the other direct.

COROLLARY 4

For all $t \geqslant 0$ and integers $k \geqslant 0$,

$$\sum_{i=k+1}^{\infty} f_{i0}(t) = F_{k0}(t) - \rho F_{k+1,0}(t).$$

First proof of corollary 4

Apply (8), (10) and corollary 2(a).

Second proof of corollary 4

Sum the equations in the recursion for the integrals of the functions in the recursion for $f_{i0}(t)$ in theorem 5.1 of [3]. □

Remark

(7) Formula (8) can be deduced from (9), corollary 2(a) and corollary 4, using the direct proof of corollary 4. This helps explain the results in section 4.2 of [2].

It is intuitively obvious that $P_{n+k,n}(t)$ and ${}^0P_{n+k,n}(t)$ converge to $U_{-k}(t)$ as $n \to \infty$. Bounds on the rate of convergence follow from (6) and theorem 1.

COROLLARY 5

For all $t \geqslant 0$ and all integers $k \geqslant 0$ and $n \geqslant 0$,

$$\text{(a)}\quad |{}^0P_{n+k,n}(t) - U_{-k}(t)| = \rho^n U_{-(2n+k)}(t) = \rho^{-(n+k)} U_{2n+k}(t) \leqslant \min\{\rho^n, \rho^{-(n+k)}\},$$

$$\text{(b)}\quad |P_{n+k,n}(t) - {}^0P_{n+k,n}(t)| = \rho^n P_{2n+k,0}(t) = \rho^{-(n+k)} P_{0,2n+k}(t) \leqslant \min\{\rho^n, \rho^{-(n+k)}\},$$

$$\text{(c)}\quad |P_{n+k,n}(t) - U_{-k}(t)| = \rho^n |P_{2n+k,0}(t) - U_{-(2n+k)}(t)| = \rho^{-(n+k)} |P_{0,2n+k}(t) - U_{2n+k}(t)| \leqslant \min\{\rho^n, \rho^{-(n+k)}\},$$

$$\text{(d)}\quad U_{-k}(t) = \sum_{j=0}^{\infty} \rho^j \left[{}^0P_{2j+k+1,1}(t)\right] = \sum_{j=0}^{\infty} \rho^j f_{2j+k+1,0}(t).$$

Proof

For (a)–(c), apply (6) and theorem 1. For (d), apply corollary 2(b) and part (a) here, letting $n \to \infty$. □

Remark

(8) Part (d) can be established by noting that the Laplace transform of $U_0(t)$ is $2\theta^2 z_1/(1-\rho z_1^2)$; see (7.5), (7.7), (2.4) and (2.5) of [2].

5. More inequalities

Recall that a function $f(x, y)$ of two real variables is TP_2 (totally positive of order 2) if $f(x_1, y_2)f(x_2, y_1) \leqslant f(x_1, y_1)f(x_2, y_2)$ whenever $x_1 < x_2$ and $y_1 < y_2$, and STP_2 (strictly TP_2) if the inequality is strict; see Karlin [12]. The following can be deduced from chapter 3 of [12].

THEOREM 2

The functions $P_{i0}(t)$, $f_{i0}(t) = b^i(t)$, $U_i(t)$, $U_{-i}(t)$ and $I_i(t)$ are STP_2 in i and t for $i \geqslant 0$ and $t \geqslant 0$.

Proof

Let A be the infinitesimal generator matrix associated with the $M/M/1$ model, with $A_{i,i+1} = \rho$, $i \geqslant 0$, and $A_{i,i-1} = 1$, $i \geqslant 1$. Let $P_\nu = I + \nu^{-1}A$ be the probability transition matrix associated with the discrete-time Markov chain (DTMC) constructed by uniformization (with ν chosen suitably large). It is easy to see that $P_\nu \equiv P_\nu(i, j)$ is TP_2 as a function of i and j for all suitably large ν; cf.

p. 113 of [12]. Let $x_i(n) = P_\nu^n(i, 0)$, i.e., the n-step probability of being in 0 starting in i for the DTMC. Since $x_0(0) = 1$ and $x_i(0) = 0$ for $i \geqslant 1$, $x_i(n+1) = \sum_{j=0}^{\infty} P_\nu(i, j) x_j(n)$ and P_ν is TP_2, we deduce that $x_i(n)$ is TP_2 in i and n. Since

$$P_{i0}(t) = \sum_{n=0}^{\infty} K(n, t) x_{i0}(n),$$

where $K(n, t)$ is the Poisson kernel $(\nu t)^n \, e^{-\nu t}/n!$, which is STP_2, $P_{i0}(t)$ is STP_2 as well. We treat $f_{i0}(t)$ essentially the same way, using the associated or lossy infinitesimal generator matrix obtained by deleting the first row and column from A; cf. section 10 of [3]. By (3) and (9), the STP_2 property for $f_{i0}(t)$ carries over to $U_i(t)$, $U_{-i}(t)$ and $I_i(t)$ too. □

We apply theorem 2 to deduce a well known ordering for the Bessel functions, which we will apply below. (See p. 151 of Magnus, Oberhettinger and Soni [14] and p. 103 of Karlin [12] for more general results.)

COROLLARY 6

$I_i(t) > I_{i+1}(t)$ for all $t > 0$ and integers i.

Proof

By theorem 2, $I_{i+1}(t)/I_i(t)$ is strictly increasing in t. By 9.7.1, p. 377, of [6], $I_{i+1}(t)/I_i(t) \to 1$ as $t \to \infty$. □

From corollary 1(d), we know that $P_{n+k,n}(t) < P_{n+k-1,n-1}(t)$ for all t when $\rho < 1$. In some cases we are able to establish a bound the other way.

THEOREM 3

Suppose that $\rho < 1$. $P_{n+k,n}(t) > \rho P_{n+k-1,n-1}(t)$ for all $t > 0$ if and only if $(1-\rho)/\sqrt{\rho} > (2n+k)/n(n+k)$ or, equivalently, $\rho < [\sqrt{(1+\beta^2)} - \beta]^2$ where $\beta = (2n+k)/2n(n+k)$. For all sufficiently large t, $P_{n+k,n}(t) < \rho P_{n+k-1,n-1}(t)$ whenever $(1-\rho)/\sqrt{\rho} < (2n+k)/n(n+k)$.

Proof

Apply theorem 1, using (a) for $P_{n+k,n}(t)$ and (b) for $P_{n+k-1,n-1}(t)$. Hence

$$P_{n+k,n}(t) - \rho P_{n+k-1,n-1}(t) = (1-\rho)^0 P_{n+k,n}(t) - \rho^n [P_{2n+k-1,0}(t) - P_{2n+k,0}(t)]$$

and, from corollaries 1(a) and 2(b),

$$\rho^{k/2}[P_{n+k,n}(t) - \rho P_{n+k-1,n-1}(t)] = \frac{(1-\rho)}{\rho^{1/2}} \sum_{i=1}^{n} \rho^{(k+2i-1)/2} f_{k+2i-1,0}(t) - \rho^{(k+2n)/2} f_{k+2n,0}(t). \tag{14}$$

By corollary 6, the Bessel functions are ordered, i.e., $I_j(t) > I_{j+1}(t)$ for all j and t, so that

$$j^{-1}\rho^{j/2} f_{j0}(t) > k^{-1}\rho^{k/2} f_{k0}(t) \quad \text{for } j<k,$$

so that $[P_{n+k,n}(t) - \rho P_{n+k-1,n-1}(t)] > 0$ provided that

$$\frac{(1-\rho)}{\rho^{1/2}} \sum_{i=1}^{n} (k+2i-1) > k+2n,$$

which reduces to the stated condition. On the other hand, when this condition does not hold, we can establish the reverse inequality asymptotically as $t\to\infty$. From theorem 3.1(b) of [2],

$$\rho^{m/2} f_{m0}(t) \sim 2m\sqrt{\rho}\, L(t,\rho) \quad \text{as } t\to\infty \tag{15}$$

for

$$L(t,\rho) = \left(\pi\rho^{3/2}t^3\right)^{-1/2} \exp(-t/\tau), \tag{16}$$

where $\tau = 1/(1-\sqrt{\rho})^2$ as in (3.1) of [2] without the time scaling. Combining (14) and (15) we get

$$P_{n+k,n}(t) - \rho P_{n+k-1,n-1}(t) \sim (1/2)\rho^{-(k-1)/2} L(t,\rho)$$
$$\times \left(\frac{(1-\rho)}{\rho^{1/2}} \left[nk+n^2\right] - (k+2n) \right) \quad \text{as } t\to\infty. \quad \square$$

We now obtain further consequences of theorem 2.

THEOREM 4

For all $t \geqslant 0$ and integers $i \geqslant 1$,

(a) $P_{i0}(t)^2 - P_{i-1,0}(t)P_{i+1,0}(t) \geqslant \rho\left(P_{i+1,0}(t)^2 - P_{i0}(t)P_{i+2,0}(t)\right) \geqslant 0,$

(b) $f_{i0}(t)^2 - f_{i-1,0}(t)f_{i+1,0}(t) \geqslant \rho\left(f_{i+1,0}(t)^2 - f_{i0}(t)f_{i+2,0}(t)\right) \geqslant 0,$

(c) $U_{-i}(t)^2 - U_{-(i-1)}(t)U_{-(i+1)}(t) \geqslant \rho\left(U_{-(i+1)}(t)^2 - U_{-i}(t)U_{-(i+2)}(t)\right) \geqslant 0.$

Proof

Note that $P_{i0}'(t)/P_{i0}(t)$ is nondecreasing in i for each t by the TP_2 property for $P_{i0}(t)$ established in theorem 2. Together with the Chapman–Kolmogorov equations

$$P_{i0}'(t) = P_{i-1,0}(t) - (1+\rho)P_{i0}(t) + \rho P_{i+1,0}(t), \quad i \geqslant 1, \tag{17}$$

this establishes the first inequality in (a). To obtain the nonnegativity in (a), note that

$$\frac{P_{i+2,0}(t)}{P_{i+1,0}(t)} = \frac{\int_0^t f_{10}(s)P_{i+1,0}(t-s)\,\mathrm{d}s}{P_{i+1,0}(t)} \leqslant \frac{\int_0^t f_{10}(s)P_{i,0}(t-s)\,\mathrm{d}s}{P_{i,0}(t)} = \frac{P_{i+1,0}(t)}{P_{i,0}(t)}, \tag{18}$$

using $P_{i+1,0}(t-s)P_{i,0}(t) \leqslant P_{i,0}(t-s)P_{i+1,0}(t)$ since $P_{i0}(t)$ is TP_2 in i and t by theorem 2. The arguments for (b) and (c) are the same, using the associated or lossy process for (b) as in theorem 2 and the unrestricted process for (c). □

As a consequence of (3) and theorem 4(c), we obtain more inequalities for the modified Bessel functions.

COROLLARY 7

For each $t > 0$, $\rho > 0$ and integer $j \geqslant 1$,

$$I_j(2\sqrt{\rho}\,t)^2 - I_{j-1}(2\sqrt{\rho}\,t)I_{j+1}(2\sqrt{\rho}\,t) \geqslant \sqrt{\rho}\left(I_{j+1}(2\sqrt{\rho}\,t)^2 - I_j(2\sqrt{\rho}\,t)I_{j+1}(2\sqrt{\rho}\,t)\right)$$
$$\geqslant 0.$$

6. Asymptotics

Another application of theorem 1 is to obtain a more elementary proof of the asymptotic behavior of $P_{ij}(t)$ as $t \to \infty$, as given on p.94 of Cohen [9]. To state the results, let $\sigma = (1-\sqrt{\rho})/\sqrt{\rho}$. Recall that $f(t) \sim g(t)$ as $t \to \infty$ if $f(t)/g(t) \to 1$ as $t \to \infty$. We only state the result for $\rho < 1$. The results for $\rho \geqslant 1$ follow in the same way.

THEOREM 5

If $\rho < 1$, then for all integers $k \geqslant 0$ and $n \geqslant 0$

(a) $${}^0P_{n+k,n}(t) \sim \frac{L(t,\rho)}{2\rho^{k/2}}n(n+k) = \frac{L(t,\rho)\sigma^2 n(n+k)}{2\sigma^2\rho^{k/2}},$$

(b) $$\rho^n P_{k0}(t) \sim (1-\rho)\rho^n + \frac{L(t,\rho)}{2\sigma^2\rho^{(k-2n)/2}}(1-k\sigma),$$

(c) $$P_{n+k,n}(t) \sim (1-\rho)\rho^n + \frac{L(t,\rho)}{2\sigma^2\rho^{k/2}}(1-n\sigma)(1-[n+k]\sigma).$$

Proof

By 9.7.1 of [6], $e^{-z}I_k(z) \sim (2\pi z)^{-1/2}$ as $z \to \infty$ for all k. For (a), apply corollary 2(b), (9) and (3). For (b), apply (3), (8), (9) and (10), as was already done in corollary 4.2.7 of [2]. For part (c), apply theorem 1 plus parts (a) and (b). □

The decompositions also help establish the asymptotic behavior with other initial distributions. As pointed out by Massey [15], the asymptotic behavior with other initial distributions is especially interesting because the relaxation time parameter (τ in (16) and theorem 5) actually depends on the initial distribution. We derive the asymptotic behavior in the case of a geometric initial distribution with parameter γ, i.e., $P(Q(0)=j)=(1-\gamma)\gamma^j$, $j \geqslant 0$. Thus, the transition function of interest is

$$P_{\gamma n}(t) = \sum_{i=0}^{\infty} (1-\gamma)\gamma^i P_{in}(t), \quad t \geqslant 0. \tag{19}$$

THEOREM 6

Suppose that $\rho < 1$ and n is an arbitrary nonnegative integer.

(a) If $\gamma < \sqrt{\rho}$, then

$$P_{\gamma n}(t) \sim (1-\rho)\rho^n + \frac{L(t,\rho)}{2\sigma^2} \frac{(1-\gamma)(\rho-\gamma)}{\left(\sqrt{\rho}-\gamma\right)^2} \rho^{n/2}(1-n\sigma).$$

(b) If $\gamma = \sqrt{\rho}$, then

$$P_{\gamma n}(t) \sim (1-\rho)\rho^n + tL(t,\rho)\frac{\rho^{(n+1)/2}}{1+\sqrt{\rho}}\left(\frac{(1-\rho)n}{2} - \rho^{(n+1)} - \frac{\sqrt{\rho}}{1+\sqrt{\rho}}\right).$$

(c) If $\sqrt{\rho} < \gamma < 1$, then

$$P_{\gamma n}(t) \sim (1-\rho)\rho^n + \left[(1-\gamma)\gamma^n - \left(1-\frac{\rho}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^n - \left(\frac{\gamma^2-\rho}{\gamma-\rho}\right)(1-\rho)\rho^n\right]$$
$$\times \exp(-t/\beta t),$$

where

$$\beta = \left(1-\sqrt{\rho}\right)^2/(1-\gamma)(1-[\rho/\gamma]). \tag{20}$$

Proof

From (11), which is valid for k negative, we obtain

$$P_{\gamma n}(t) = {}^0P_{\gamma n}(t) + f_{\gamma 0}(t) * P_{0n}(t), \tag{21}$$

where $f_{\gamma 0}(t)$ is the density of the first passage time to 0 starting with initial distribution γ (including an atom of size $(1-\gamma)$ at 0). Let $F_{\gamma 0}(t)$ be the associated cdf and let $F_{\gamma 0}^c(t)=1-F_{\gamma 0}(t)$, and similarly for $F_{n0}(t)$. Then, from (8), we obtain

$$f_{\gamma 0}(t) * P_{0n}(t)=(1-\rho)\rho^n-(1-\rho)\rho^n F_{\gamma 0}^c(t)+\rho^{n+1} f_{\gamma 0}(t) * F_{n+1,0}^c(t) \\ -\rho^n f_{\gamma 0}(t) * F_{n,0}^c(t).$$

For the zero-avoiding component of (21), write

$${}^0P_{\gamma n}(t)=\rho^n(1-\gamma)\sum_{i=0}^{n-1}\left(\frac{\gamma}{\rho}\right)^i {}^0P_{in}(t)+(1-\gamma)\sum_{i=n}^{\infty}(\gamma^i){}^0P_{in}(t),$$

from which we obtain (by applying corollary 2(b) and transform manipulation)

$${}^0P_{\gamma n}(t)=\frac{(1+\rho)^2\gamma^n}{2}\sum_{i=0}^{n-1}\left(\frac{\rho}{\gamma}\right)^i f_{i0}(t) * h_\gamma(t), \tag{23}$$

where $h_\gamma(t)$ is the density of the conditional time to emptiness starting with the geometric distribution with parameter γ, conditional on being strictly positive; see corollary 3.1.3 of [1] and theorems 3.3 and 5.1 of [2]. Note that the Laplace transforms of $f_{i0}(t)$, $f_{\gamma 0}(t)$ and $h_\gamma(t)$ are $\hat{f}_{i0}(s)=z_1^i$, $\hat{f}_{\gamma 0}(s)=(1-\gamma)/(1-\gamma z_1)$ and $\hat{h}_\gamma(s)=(1-\gamma)z_1/(1-\gamma z_1)$ for $z_1\equiv z_1(s)$ being the relevant root of the quadratic equation as in (2.4) of [2] without time scaling. To determine the asymptotic behavior of $f_{\gamma 0}(t)$ and $P_{\gamma n}(t)$ as $t\to\infty$, we must identify the rightmost singularities of the transform $\hat{f}_{\gamma 0}(s)$ and $\hat{P}_{\gamma n}(s)$. This involves solving the equation $1-\gamma z_1(s)=0$. There is a negative root to this equation if and only if $\gamma>\rho$, in which case it is given by $s=-1/\beta\tau$ for β in (20). Hence, $\hat{f}_{\gamma 0}(s)$ and $\hat{P}_{\gamma n}(s)$ have simple pole singularities at $s=-1/\beta\tau$. However, these functions (since they involve $z_1(s)$) also have branch point singularities at $s=-1/\tau$. Hence, the pole is the right-most singularity when $\beta>1$ or, equivalently, when $\gamma>\sqrt{\rho}$. For $\gamma>\sqrt{\rho}$ and $\gamma<\sqrt{\rho}$, we obtain the asymptotic expansions by applying Heaviside's theorem on p. 254 of Doetsch [11]. For $\gamma=\sqrt{\rho}$, we see that

$$\hat{f}_{\gamma 0}(s)=\left(\frac{1+\sqrt{\rho}}{2}\right)\left(\frac{\tau_1 s+1}{\tau_0 s+1}\right)^{1/2}+\frac{1-\sqrt{\rho}}{2},$$

from which we directly see that

$$f_{\gamma 0}(t)\sim\frac{2\rho t}{1+\sqrt{\rho}}L(t,\rho)$$

and

$$f_{\gamma 0}(t) * f_{n,0}(t)\sim\frac{2\rho t L(t,\rho)}{\rho^{n/2}(1+\sqrt{\rho})}$$

as needed for (b). □

Remarks

(9) It is instructive to see what happens if we try to apply theorem 5(c), acting as if the limits in the summation in (19) and $t \to \infty$ can be exchanged. We obtain

$$P_{\gamma n}(t) \sim (1-\rho)\rho^n + \frac{L(t, \rho)\rho^{n/2}}{2\sigma^2}(n\sigma - 1)(1-\gamma)$$
$$\times \left[\sigma \sum_{i=0}^{\infty} i \left(\frac{\gamma}{\sqrt{\rho}} \right)^i - \sum_{i=0}^{\infty} \left(\frac{\gamma}{\sqrt{\rho}} \right)^i \right],$$

from which the obtain the result in theorem 6(a) when $\gamma < \sqrt{\rho}$, but this method does not work for $\gamma \geqslant \sqrt{\rho}$.

(10) Note that the result in theorem 6(b) is not obtained by taking the limit as $\gamma \to \sqrt{\rho}$ in parts (a) and (c).

(11) Massey [15] obtained the results in theorem 6(a) and (c), but there seems to be an error in his expression for the constant multiplying the exponential in his version of (c); see his (20).

(12) Analogs of theorem 6 for other initial distributions do not seem nearly so easy to determine by our approach. See [15] for an alternate approach.

(13) Consistent with our previous experience, numerical comparisons indicate that the pure-exponential limit in theorem 6(c) is a much more accurate approximation for typical times t than the limits in the other cases.

(14) (added in proof) Previous work related to theorem 6, including (20), appears in J.P.C. Blanc and E.A. van Doorn, Relaxation times for queueing systems, in: *Proc. CWI Symp. on Mathematics and Computer Science*, eds. J.W. de Bakker, M. Hazewinkel and J.K. Lenstra (North-Holland, Amsterdam, 1984) pp. 139–162.

References

[1] J. Abate and W. Whitt, Transient behavior of the $M/M/1$ queue: starting at the origin, Queueing Systems 2 (1987) 41–65.

[2] J. Abate and W. Whitt, Transient behavior of the $M/M/1$ queue via Laplace transforms, Adv. Appl. Prob. 20 (1988) 145–178.

[3] J. Abate and W. Whitt, Simple spectral representations for the $M/M/1$ queue, Queueing Systems 3 (1988) 321–346.

[4] J. Abate and W. Whitt, Calculating time-dependent performance measures for the $M/M/1$ queue, IEEE Trans. Commun. COM-37 (1989) 1102–1104.

[5] J. Abate and W. Whitt, In search of the camel: the shape of the $M/M/1$ transition function, in preparation.

[6] M. Abramowitz and I.A. Stegun (eds.), *Handbook of Mathematical Functions* (Dover, New York, 1972).

[7] F. Baccelli and W.A. Massey, A sample path analysis of the $M/M/1$ queue, J. Appl. Prob. 26 (1989) 418–422.

[8] N.T.J. Bailey, A continuous time treatment of a simple queue using generating functions, J. Roy. Statist. Soc. B16 (1954) 288–291.
[9] J.W. Cohen, *The Single Server Queue*, 2nd ed. (North-Holland, Amsterdam, 1982).
[10] B. Conolly, *Lecture Notes on Queueing Systems* (Wiley, New York, 1975).
[11] G. Doetsch, *Introduction to the Theory and Application of Laplace Transformation* (Springer, New York, 1974).
[12] S. Karlin, *Total Positivity* (Stanford University Press, Stanford, CA, 1968).
[13] S. Karlin and J.L. McGregor, A characterization of birth and death processes, Proc. Nat. Acad. Sci. USA 45 (1959) 375–379.
[14] W. Magnus, F. Oberhettinger and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 3rd ed. (Springer, New York, 1966).
[15] W.A. Massey, The $M/M/1$ queue relaxation parameter depends on its initial distribution, AT&T Bell Laboratories, Murray Hill, NJ (1989).
[16] N.U. Prabhu, *Queues and Inventories* (Wiley, New York, 1965).
[17] N.U. Prabhu, *Stochastic Storage Processes* (Springer, New York, 1980).