Joint distribution of sojourn time and queue length in the M/G/1 queue with (in)finite capacity

O.J. BOXMA
Mathematical Institute, University of Utrecht, Utrecht, Netherlands

Received September 1982
Revised December 1982

For the M/G/1 queue we study the joint distribution of the number of customers x present immediately before an arrival epoch and of the residual service time ζ of the customer in service at this epoch. The correlation coefficient $\rho(x, \zeta)$ is shown to be positive (negative) when the service time distribution is DFR (IFR). The result for the joint distribution of x and ζ leads to the joint distribution of x, of the sojourn time s of the arriving customer and of the number of customers z left behind by this customer at his departure. $\rho(x, s)$, $\rho(z, s)$ and $\rho(x, z)$ are shown to be positive; $\rho(x, s)$ and $\rho(z, s)$ are compared in some detail.

Subsequently the M/G/1 queue with finite capacity is considered; the joint distributions of x and ζ and of x and s are derived. These results may be used to study the cycle time distribution in a two-stage cyclic queue.

1. Introduction

In many studies of queuing models problems of the following type occur: determine the distribution of the residual service time(s) of the customer(s) in service at an arrival or departure epoch, if it is known that k customers are present at such an epoch. For example in some control problems of M/G/1 type queues, certain actions taken at arrival instants of customers depend on both the queue length at such an instant, and the residual service time of the customer in service—and clearly these two variables are not independent.

Quite a few queuing approximations are based on approximation assumptions concerning the conditional residual service time distribution, cf. Nozaki and Ross [10], Tijms and Van Hoorn [11]. In a recent study [3] we have come across the problem of determining the joint distribution of the residual service time and the number of customers present immediately before an arrival in the M/G/1 queue with finite capacity N. In this case an exact analysis turns out to be possible; it is presented here.

The case $N = \infty$ (M/G/1 queue) has already been discussed in an early paper by Wishart [13], while many results can also be found in Cohen [4]. Recently there has been a resurge of interest in this case. In the appendix of a 1979 Tel Aviv University report of Mandelbaum and Yechiali (it will be published [8] without this appendix) the conditional mean residual service time given the queue length at arrival epochs was derived. The same result (at an arbitrary epoch) was obtained by Fakinos [5]. Asmussen [1] derives expressions for the densities of residual and past service time at an arbitrary epoch given the queue length at such an epoch. He, too, finds the same expression for conditional mean residual service time.

In the present study we extend these results in the following ways: We derive for both the M/G/1 and M/G/1/N model an expression for covariance and correlation coefficient of residual service time and queue length. Also, we obtain expressions for the joint distribution of queue length immediately before an arrival epoch, sojourn (response) time and (for the case $N = \infty$) queue length immediately after a departure epoch.

North-Holland
European Journal of Operational Research 16 (1984) 246–256

The description of the M/G/1 model is specified by stating that the arrival intensity is α^{-1} and that the service times are independent, identically distributed non-negative stochastic variables τ_i, $i = 1, 2, \ldots$ with distribution $B(\ )$ with Laplace–Stieltjes Transform (LST) $\beta(\ )$, finite first moment β and nth moment β_n, $n = 2, 3, \ldots$; we define $a = \beta/\alpha$.

In the case that the system has a finite capacity N (i.e., the system can contain at most N customers, including the one in service), it is assumed that an arriving customer who finds N customers present leaves immediately, never to return. In the present study such an event will not be considered as an arrival.

In the following we denote by ζ, x, s and z stochastic variables (s.v.) with distributions the limiting distributions for $n \to \infty$ (if these exist) of the residual service time ζ_n at the nth arrival epoch after $t = 0$, the number of customers x_n present immediately before this epoch, the sojourn time s_n of the arriving customer and the number of customers z_n left behind by this customer at his departure, respectively.

Of course, in the M/G/1 queue the limiting distributions of x_n, s_n and z_n exist iff $a < 1$, whereas in the M/G/1/N queue these limiting distributions exist irrespective of the value of a (cf. [4]).

The organization of this article is as follows. In Section 2 we study the joint distribution of x and ζ in the M/G/1 queue. In particular it is shown that the correlation coefficient $\rho(x, \zeta)$ is positive when $B(\ )$ is a DFR (decreasing failure rate) distribution, and negative when $B(\ )$ is an IFR (increasing failure rate) distribution. The joint distribution of x, s and z in the M/G/1 queue is studied in Section 3. It is shown that the joint distributions of x and s and of z and s coincide iff G ≡ M (M/M/1 queue). Some results on the (light and heavy traffic) behaviour of $\rho(x, s)$, $\rho(z, s)$ and $\rho(x, z)$ may add to our insight into the behaviour of the M/G/1 queue.

In Section 4 we turn to the M/G/1/N queue; we derive the joint distributions of x and ζ and of x and s. These results are used in [3] to study the joint distribution of the response times in the two queues of a two-stage cyclic queuing system.

2. Joint distribution of x and ζ in the M/G/1 queue

Consider the M/G/1 queue with infinite capacity, and suppose that $a = \beta/\alpha < 1$. Let

$$G(j, v)\,\mathrm{d}v \stackrel{\text{def}}{=} \Pr\{x = j,\ v \leqslant \zeta < v + \mathrm{d}v\}, \quad j = 1, 2, \ldots, v > 0, \tag{2.1}$$

$$Q(s, \rho) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} s^j \int_{v=0}^{\infty} \mathrm{e}^{-\rho v} G(j, v)\,\mathrm{d}v, \quad |s| \leqslant 1,\ \mathrm{Re}\,\rho \geqslant 0. \tag{2.2}$$

Theorem 2.1.

$$Q(s, \rho) = E\left[s^x \mathrm{e}^{-\rho\zeta}(x > 0)\right] = \frac{(1-a)s}{\beta\left(\frac{1-s}{\alpha}\right) - s}(1-s)\frac{\beta(\rho) - \beta\left(\frac{1-s}{\alpha}\right)}{1 - \alpha\rho - s}, \tag{2.3}$$

((·) *denotes indicator function*), $|s| \leqslant 1$, Re $\rho \geqslant 0$.

Proof. See Wishart [13]. In [12] Wishart derives the same formula for the M/G/1 queue with LCFS service discipline, considered at an arbitrary epoch.

Remark 2.1. Expression (2.3) completely agrees with the transform of the joint distribution of queue length and residual service time at an *arbitrary* epoch, cf. (II.4.88) in [4]. This could have been predicted because of the memoryless property of the interarrival time distribution.

Remark 2.2. Asmussen [1] yields expressions for the densities of ζ and of η (past service time) at an arbitrary epoch under the condition that the queue length at this epoch is equal to n. Because of Remark

2.1 it is obvious that his expressions for the conditional moments of ζ agree with those when the queue length at an arrival epoch is given. In particular

$$E[\zeta \mid x = n] = \alpha(1 - a)\frac{\Pr\{x > n\}}{\Pr\{x = n\}} \tag{2.4}$$

(see also Mandelbaum and Yechiali [8] and Fakinos [5]).

Remark 2.3. From (2.3) it is easy to derive the Pollaczek–Khinchine formula for the LST of the distribution of the waiting time w in an M/G/1 queue:

$$\begin{aligned} E[\mathrm{e}^{-\rho w}] &= \Pr\{w = 0\} + \sum_{k=1}^{\infty} \Pr\{x = k\} E[\mathrm{e}^{-\rho w} \mid x = k] \\ &= (1 - a) + \sum_{k=1}^{\infty} \beta^{k-1}(\rho) E[\mathrm{e}^{-\rho\zeta}(x = k)] \\ &= (1 - a) + Q(\beta(\rho), \rho)/\beta(\rho) = (1 - a)\frac{\alpha\rho}{\alpha\rho + \beta(\rho) - 1}, \quad \operatorname{Re}\rho \geqslant 0. \end{aligned} \tag{2.5}$$

Remark 2.4. Note that $Q(s, 1/\alpha)$ has the following meaning:

$$\begin{aligned} Q\left(s, \frac{1}{\alpha}\right) &= \sum_{j=1}^{\infty} s^j \int_0^{\infty} \mathrm{e}^{-t/\alpha}\, \mathrm{d}\Pr\{x = j, \zeta < t\} = \sum_{j=1}^{\infty} s^j \Pr\{x = j, \zeta < \sigma\} \\ &= \sum_{j=1}^{\infty} s^j \Pr\{x = j, \tilde{z} = j\}, \end{aligned}$$

with σ an interarrival time and $\tilde{z}$ the queue length immediately after the first departure following the arrival epoch we are considering. So from (2.3):

$$\sum_{j=1}^{\infty} s^j \Pr\{x = j, \tilde{z} = j\} = (1 - a)(1 - s)\frac{\beta\left(\frac{1-s}{\alpha}\right) - \beta\left(\frac{1}{\alpha}\right)}{\beta\left(\frac{1-s}{\alpha}\right) - s}, \quad |s| \leqslant 1. \tag{2.6}$$

Theorem 2.2.

$$E[x\zeta] = \frac{\beta_3}{6\alpha^2} + \frac{\beta_2^2}{4\alpha^3(1 - a)} + \frac{\beta_2}{2\alpha}, \tag{2.7}$$

$$\operatorname{cov}(x, \zeta) = a^2\beta\left[\frac{\beta_3}{6\beta^3} - \left(\frac{\beta_2}{2\beta^2}\right)^2\right], \tag{2.8}$$

$$\begin{aligned} \rho(x, \zeta) = a^2\beta\left[\frac{\beta_3}{6\beta^3} - \left(\frac{\beta_2}{2\beta^2}\right)^2\right]\Bigg[&\left\{\frac{\beta_3}{3\beta} - \left(\frac{\beta_2}{2\beta}\right)^2\right\}^{1/2} \\ &\times\left\{\frac{\beta_2}{\alpha^2} + \frac{\beta_2}{2\alpha^2(1 - a)} + a - a^2 + \frac{\beta_2^2}{4\alpha^4(1 - a)^2} + \frac{\beta_3}{3\alpha^3(1 - a)}\right\}^{1/2}\Bigg]^{-1}. \end{aligned} \tag{2.9}$$

Proof. (2.7) follows from (2.3) after a lengthy but straightforward calculation. (2.8) and (2.9) follow from

(2.7) and the well-known relations (which also can be derived from (2.3)):

$$E[\zeta]=\frac{\beta_2}{2\beta},\qquad E[\zeta^2]=\frac{\beta_3}{3\beta},\tag{2.10}$$

$$E[x]=\frac{\beta_2}{2\alpha^2(1-a)}+a,\qquad E[x^2]=\frac{\beta_3}{3\alpha^3(1-a)}+\frac{\beta_2^2}{2\alpha^4(1-a)^2}+\frac{3\beta_2}{2\alpha^2(1-a)}+a.\tag{2.11}$$

If $\log(1-B(x))$ is concave (convex) in $x \geqslant 0$, then $B(\ )$ is said to have an increasing (decreasing) failure rate (IFR, DFR). Roughly speaking, if B is IFR then the fact that the past service time η at an arrival epoch is relatively large implies that it is likely that ζ will be relatively small, and vice versa. Hence the following corollary is not surprising:

Corollary 2.1.

$\rho(x,\zeta)>0$ *if* $B(\ )$ *is* DFR;

$\rho(x,\zeta)<0$ *if* $B(\ )$ *is* IFR;

$\rho(x,\zeta)=0$ *if* $B(x)=1-e^{-x/\beta}$.

Proof. We have to consider the sign of $\beta_3/6\beta^3-(\beta_2/2\beta^2)^2$. According to [2], see also [9, pp. 362/494], $\{\log E[\tau^r]\}/\Gamma(r+1)$ is concave (convex) if τ has a DFR (IFR) distribution. Hence $\log\beta+\log(\beta_3/6)\leqslant$ $(\geqslant)2\log(\beta_2/2)$ if $B(\ )$ is IFR (DFR) and the result follows.

An obvious by-product of Theorem 2.2 is:

Corollary 2.2.

$$\lim_{a\to 1}\frac{\rho(x,\zeta)}{1-a}=\frac{\beta\left[\dfrac{\beta_3}{6\beta^3}-\left(\dfrac{\beta_2}{2\beta^2}\right)^2\right]}{\left\{\dfrac{\beta_3}{3\beta}-\left(\dfrac{\beta_2}{2\beta}\right)^2\right\}^{1/2}\dfrac{\beta_2}{2\beta^2}},\tag{2.12}$$

$$\lim_{\alpha\to\infty}\alpha^{3/2}\rho(x,\zeta)=\frac{\dfrac{\beta_3}{6}-\dfrac{\beta_2^2}{4\beta}}{\left\{\dfrac{\beta_3}{3\beta}-\left(\dfrac{\beta_2}{2\beta}\right)^2\right\}^{1/2}\beta^{1/2}}.\tag{2.13}$$

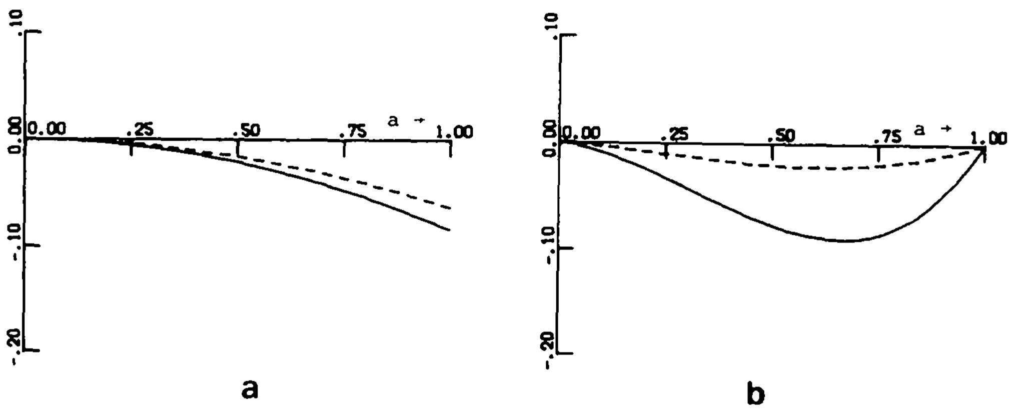

Fig. 1. cov(x,ζ) (a) and $\rho(x,\zeta)$ (b) for the M/E$_2$/1 (— — —) and M/D/1 (———) queue ($\beta=1$ fixed).

Eqs. (2.12) and (2.13) confirm our intuitive feeling that both in heavy traffic ($a \to 1$) and in light traffic ($\alpha \to \infty$, β fixed) the length of the past service time should hardly influence the queue length. See also Fig. 1, where $\text{cov}(x, \zeta)$ and $\rho(x, \zeta)$ are plotted against the traffic intensity a for the $M/E_2/1$ and $M/D/1$ case.

3. Joint distribution of x, s and z in the M/G/1 queue

Theorem 3.1. *For $a < 1$, the joint distribution of x, the queue length immediately before an arrival, s, the sojourn time of the arriving customer, and z, the queue length immediately after his departure, is given by*

$$E[s^x e^{-\rho s} r^z] = (1-a)\beta\left(\rho + \frac{1-r}{\alpha}\right) + Q\left(s\beta\left(\rho + \frac{1-r}{\alpha}\right), \rho + \frac{1-r}{\alpha}\right), \quad |s| \leqslant 1, \operatorname{Re}\rho \geqslant 0, |r| \leqslant 1. \tag{3.1}$$

Proof. For $|s| \leqslant 1$, $\operatorname{Re}\rho \geqslant 0$, $|r| \leqslant 1$,

$$\begin{aligned}
&E[s^x e^{-\rho s} r^z] \\
&= \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} r^j \int_0^\infty e^{-\rho t} \Pr\{z = j \mid x = k, s = t\} \, d_t \Pr\{x = k, s < t\} \\
&\overset{(1)}{=} \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} r^j \int_0^\infty e^{-\rho t} \frac{(t/\alpha)^j}{j!} e^{-t/\alpha} \, d_t \Pr\{s < t \mid x = k\} \Pr\{x = k\} \\
&= \Pr\{x = 0\} \int_0^\infty e^{-(\rho + (1-r)/\alpha)t} \, dB(t) \\
&\quad + \sum_{k=1}^{\infty} s^k \Pr\{x = k\} \int_0^\infty e^{-(\rho + (1-r)/\alpha)t} \, d_t \Pr\{\tau_1 + \cdots + \tau_x + \zeta < t \mid x = k\} \\
&\overset{(2)}{=} (1-a)\beta\left(\rho + \frac{1-r}{\alpha}\right) + \sum_{k=1}^{\infty} s^k \beta^k\left(\rho + \frac{1-r}{\alpha}\right) E\left[e^{-(\rho + (1-r)/\alpha)\zeta}(x = k)\right];
\end{aligned} \tag{3.2}$$

Step (1) holds because z is independent of x, given s, while step (2) holds because ζ and $\tau_1, \ldots, \tau_x$ are independent given x. Now (3.1) follows from (3.2) and (2.2).

Theorem 3.2.

$$E[xs] = E[x\zeta] + \beta E x^2 = \frac{\beta_3}{6\alpha^2} \frac{1+a}{1-a} + \frac{\beta_2^2}{4\alpha^3} \frac{1+a}{(1-a)^2} + \frac{\beta_2}{2\alpha(1-a)}(2a+1) + \beta a; \tag{3.3}$$

$$E[zs] = \frac{1}{\alpha} E[s^2] = \frac{\beta_3}{6\alpha^2} \frac{2}{1-a} + \frac{\beta_2^2}{4\alpha^3} \frac{2}{(1-a)^2} + \frac{\beta_2}{\alpha(1-a)}; \tag{3.4}$$

$$E[xz] = \frac{1}{\alpha} E[xs] = \frac{\beta_3}{6\alpha^3} \frac{1+a}{1-a} + \frac{\beta_2^2}{4\alpha^4} \frac{1+a}{(1-a)^2} + \frac{\beta_2}{2\alpha^2(1-a)}(2a+1) + a^2; \tag{3.5}$$

$$\text{cov}(x, s) = \alpha\, \text{cov}(x, z) = \frac{\beta_3}{6\alpha^2} \frac{1+a}{1-a} + \frac{\beta_2^2}{4\alpha^3(1-a)^2} a + \frac{\beta_2}{2\alpha(1-a)}; \tag{3.6}$$

$$\text{cov}(z, s) = \frac{\beta_3}{3\alpha^2(1-a)} + \frac{\beta_2^2}{4\alpha^3(1-a)^2} + \frac{\beta_2 - \beta^2}{\alpha}; \tag{3.7}$$

$$\rho(x, s) = \text{cov}(x, s) \Big/ \left\{ \frac{1}{\alpha^2} (\sigma^2(s))^2 + Ex\sigma^2(s) \right\}^{1/2}; \tag{3.8}$$

$$\rho(z, s) = \mathrm{cov}(z, s) / \left\{ \frac{1}{\alpha^2} (\sigma^2(s))^2 + Ex\sigma^2(s) \right\}^{1/2}; \tag{3.9}$$

$$\rho(x, z) = \mathrm{cov}(x, z) / \sigma^2(x); \tag{3.10}$$

where

$$\sigma^2(x) = \frac{\beta_3}{3\alpha^3(1-a)} + \frac{\beta_2^2}{4\alpha^4(1-a)^2} + \frac{\beta_2 - \beta^2}{\alpha^2} + \frac{\beta_2}{2\alpha^2(1-a)} + a,$$

$$\sigma^2(s) = \frac{\beta_3}{3\alpha(1-a)} + \frac{\beta_2^2}{4\alpha^2(1-a)^2} + \beta_2 - \beta^2.$$

Proof. (3.3) follows either from (3.1) after a lengthy calculation, or from (2.7), (2.11) and the following reasoning:

$$\begin{aligned} E[xs] &= E[x(\zeta + \tau_1 + \cdots + \tau_x)] \\ &= E[x\zeta] + E_x[E[(x(\tau_1 + \cdots + \tau_x))|x]] = E[x\zeta] + \beta E[x^2]. \end{aligned} \tag{3.11}$$

(3.4) follows easily from (3.1) by observing that $E[r^z e^{-\rho s}] = E[e^{-(\rho + (1-r)/\alpha)s}]$. (3.5) follows from (3.1) and (3.3) by comparing the expressions for $E[s^x e^{-\rho s}]$ and $E[s^x r^z]$. Note that the first part of (3.5) (and of (3.6)) displays a kind of Little formula. (3.6)–(3.10) follow from (3.3)–(3.5), application of Little's formula $Es = (1/\alpha)Ex (= (1/\alpha)Ez)$ and known formulae for the moments of x and s (see also (2.11)).

An obvious consequence of Theorem 3.2 is:

Corollary 3.1.

$$\rho(x, s) \geq 0, \quad \rho(z, s) \geq 0, \quad \rho(x, z) \geq 0; \tag{3.12}$$

$$\lim_{a \to 1} \rho(x, s) = \lim_{a \to 1} \rho(z, s) = \lim_{a \to 1} \rho(x, z) = 1; \tag{3.13}$$

$$\lim_{\alpha \to \infty} \sqrt{\alpha}\, \rho(x, s) = \frac{\beta_2}{2\beta^{1/2}(\beta_2 - \beta^2)^{1/2}}, \quad \beta_2 \neq \beta^2,$$

$$\lim_{\alpha \to \infty} \rho(x, s) = \tfrac{1}{2}\sqrt{3}, \quad \beta_2 = \beta^2; \tag{3.14}$$

$$\lim_{\alpha \to \infty} \sqrt{\alpha}\, \rho(z, s) = \{(\beta_2 - \beta^2)/\beta\}^{1/2}, \quad \beta_2 \neq \beta^2;$$

$$\lim_{\alpha \to \infty} \alpha\rho(z, s) = (\beta_3/3\beta)^{1/2}, \quad \beta_2 = \beta^2; \tag{3.15}$$

$$\lim_{\alpha \to \infty} \alpha\rho(x, z) = \beta_2/2\beta. \tag{3.16}$$

Remark 3.1. The fact that the queue length process in an M/M/1 queue is a birth and death process, and hence reversible (cf. Kelly [6]) suggests that in the M/M/1 case $E[s^x e^{-\rho s}] = E[s^z e^{-\rho s}]$. Indeed from (3.1) after some arithmetic

$$E[s^x e^{-\rho s}] = E[e^{-(\rho + (1-s)/\alpha)s}] = E[s^z e^{-\rho s}] = \frac{1-a}{1 + \beta\rho - as}, \quad \mathrm{Re}\, \rho \geq 0, |s| \leq 1.$$

However, $E[s^x e^{-\rho s}] \neq E[s^z e^{-\rho s}]$ if G ≢ M, although according to (3.13), (3.14) and (3.15) $\rho(x, s)$ and $\rho(z, s)$ have in general (G ≠ D) about the same behaviour, in light and heavy traffic situations. See also Fig. 2, where $\rho(x, s)$ and $\rho(z, s)$ are plotted against the traffic intensity a for the M/E_2/1 and M/D/1 case. Note that $\rho(x, s) < (>) \rho(z, s)$ when $\beta_2 > (<) 2\beta^2$.

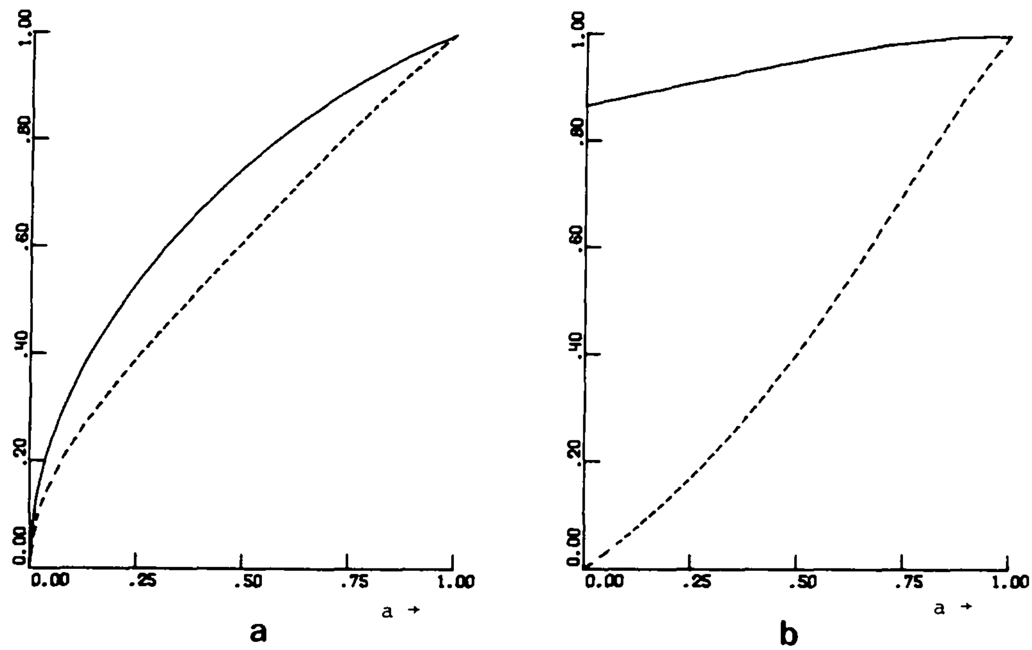

Fig. 2. Comparison of $\rho(x,s)$ and $\rho(z,s)$ for the $M/E_2/1$ and M/D/1 queue ($\beta = 1$ fixed). (a) $\rho(x,s)$ (———) and $\rho(z,s)$ (— — —) for $M/E_2/1$; (b) idem for M/D/1.

4. Joint distribution of x and ζ and of x and s in the $M/G/1/N$ queue

Considering the M/G/1 queue with finite capacity N (waiting places + service place) let

$$P(s,\rho) \overset{\text{def}}{=} \sum_{j=1}^{N-1} s^j \int_0^\infty e^{-\rho v} \, d_v \Pr\{x=j, \zeta < v\}, \quad \operatorname{Re} \rho \geqslant 0. \tag{4.1}$$

Further let D_ω be a circle with center at 0 and radius ω, $|\omega| < \mu_0$ with μ_0 the zero of $p - \beta((1-p)/\alpha)$ which is smallest in absolute value.

Theorem 4.1.

$$P(s,\rho) = \frac{1}{C}\frac{1}{2\pi i}\int_{D_\omega} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \frac{\beta\left(\frac{1-\omega}{\alpha}\right) - \beta(\rho)}{\alpha\rho + \omega - 1}\left[1 - \left(\frac{s}{\omega}\right)^{N-1}\right]\frac{s}{s-\omega}\, d\omega, \tag{4.2}$$

$$\operatorname{Re}\rho \geqslant 0, \; |\omega| < \mu_0,$$

with

$$C = \frac{1}{2\pi i}\int_{D_\omega} \frac{1}{\beta\left(\frac{1-\omega}{\alpha}\right) - \omega}\frac{d\omega}{\omega^{N-1}}, \quad |\omega| < \mu_0. \tag{4.3}$$

Proof. Theorem 4.1 may be proved similarly as Theorem 2.1, with a few changes incorporating the fact that $N < \infty$. However, we shall omit this lengthy derivation here and take our recourse to the kind of reasoning mentioned in Remark 2.1, based on the memoryless property of the interarrival time distribution.

Cohen [4, Section III.6.3] considers the M/G/1/N queue, with $\boldsymbol{x}_t$, $\boldsymbol{\eta}_t$ denoting queue length and past service time at t, $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{\eta}}$ denoting s.v. with distributions the limiting distributions of $\boldsymbol{x}_t$, $\boldsymbol{\eta}_t$ for $t \to \infty$, and with

$$R_0 \overset{\text{def}}{=} \Pr\{\hat{\boldsymbol{x}} = 0\},$$

$$R_j(\eta)\,\mathrm{d}\eta \overset{\text{def}}{=} \Pr\{\hat{\boldsymbol{x}} = j, \eta \leqslant \hat{\boldsymbol{\eta}} < \eta + \mathrm{d}\eta\}, \quad j = 1,\dots,N, \eta > 0,$$

$$R_j \overset{\text{def}}{=} \Pr\{\hat{\boldsymbol{x}} = j\} = \int_0^\infty R_j(\eta)\,\mathrm{d}\eta. \tag{4.4}$$

He proves that the following holds:

$$\Pr\{\boldsymbol{x} = j\} = \frac{R_j}{1 - R_N}, \quad j = 0, 1,\dots, N-1. \tag{4.5}$$

Here, for $|\omega| < \mu_0$,

$$R_0 = -\frac{\alpha}{D} \overset{\text{def}}{=} \left[1 + \frac{a}{2\pi\mathrm{i}} \int_{D_\omega} \frac{\mathrm{d}\omega}{\omega^{N-1}} \frac{1}{\beta\left(\frac{1-\omega}{\alpha}\right) - \omega}\right]^{-1}. \tag{4.6}$$

$$R_j(\eta) = \frac{1 - B(\eta)}{2\pi\mathrm{i} D} \int_{D_\omega} \frac{\mathrm{d}\omega}{\omega^j} \frac{(1-\omega)\mathrm{e}^{-\eta(1-\omega)/\alpha}}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)}, \quad j = 1,\dots, N-1, \eta > 0, \tag{4.7}$$

$$R_N(\eta) = \frac{1 - B(\eta)}{2\pi\mathrm{i} D} \int_{D_\omega} \frac{\mathrm{d}\omega}{\omega^{N-1}} \frac{1 - \mathrm{e}^{-\eta(1-\omega)/\alpha}}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)}, \quad \eta > 0, \tag{4.8}$$

$$R_j = -\frac{\alpha}{2\pi\mathrm{i} D} \int_{D_\omega} \frac{\mathrm{d}\omega}{\omega^j} \frac{1 - \beta\left(\frac{1-\omega}{\alpha}\right)}{\beta\left(\frac{1-\omega}{\alpha}\right) - \omega}, \quad j = 1,\dots, N-1, \tag{4.9}$$

$$R_N = -\frac{\alpha}{2\pi\mathrm{i} D} \int_{D_\omega} \frac{\mathrm{d}\omega}{\omega^{N-1}} \frac{1}{\beta\left(\frac{1-\omega}{\alpha}\right) - \omega} \left\{a - \frac{1 - \beta\left(\frac{1-\omega}{\alpha}\right)}{1-\omega}\right\}. \tag{4.10}$$

Note that C, defined in (4.3), is related to the above expressions in the following way:

$$\frac{1}{C} = -\frac{\alpha}{D} \frac{1}{1 - R_N} = \frac{R_0}{1 - R_N} = \Pr\{\boldsymbol{x} = 0\}. \tag{4.11}$$

In his derivation of the LST of the waiting time distribution, Cohen subsequently generalizes (4.4), without explicitly writing this down, to

$$\Pr\{\boldsymbol{x} = j, \eta \leqslant \boldsymbol{\eta} < \eta + \mathrm{d}\eta\} = \Pr\{\hat{\boldsymbol{x}} = j, \eta \leqslant \hat{\boldsymbol{\eta}} < \eta + \mathrm{d}\eta\}/(1 - R_N) = \frac{R_j(\eta)}{1 - R_N}\,\mathrm{d}\eta,$$
$$\eta > 0, j = 1,\dots, N-1. \tag{4.12}$$

Using (4.12) and (4.7) we can write for $\operatorname{Re}\rho \geqslant 0$, $j = 1,\dots, N-1$,

$$E\left[\mathrm{e}^{-\rho\boldsymbol{\zeta}}(\boldsymbol{x} = j)\right] = \int_{\eta=0}^\infty \frac{R_j(\eta)}{1 - R_N} E\left[\mathrm{e}^{-\rho\boldsymbol{\zeta}} \mid \boldsymbol{x} = j, \boldsymbol{\eta} = \eta\right] \mathrm{d}\eta =$$

$$= \int_{\eta=0}^{\infty} \frac{R_j(\eta)}{1-R_N} \int_{t=0}^{\infty} e^{-\rho t} d_t \left\{ \frac{B(t+\eta)-B(\eta)}{1-B(\eta)} \right\} d\eta$$
$$= \frac{1}{1-R_N} \frac{1}{2\pi i D} \int_{D_\omega} \frac{d\omega}{\omega^j} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \int_{\eta=0}^{\infty} e^{-\eta(1-\omega)/\alpha} \int_{t=0}^{\infty} e^{-\rho t} d_t B(t+\eta)\, d\eta$$
$$= \frac{\alpha/D}{(1-R_N)} \frac{1}{2\pi i} \int_{D_\omega} \frac{d\omega}{\omega^j} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \frac{\beta\left(\frac{1-\omega}{\alpha}\right) - \beta(\rho)}{\alpha\rho + \omega - 1}. \tag{4.13}$$

From (4.11) and (4.13) the statement of the theorem follows.

Theorem 4.2. *For the* M/G/1/*N queue,*

$$E[s^x e^{-\rho s}] = \frac{1}{C}\beta(\rho) + P(s\beta(\rho), \rho), \quad \text{Re}\,\rho \geqslant 0. \tag{4.14}$$

Proof. The proof of (4.14) follows almost the same lines as the proof of (3.1); using the fact that ζ and $\tau_1, \ldots, \tau_x$ are independent given $x = j$, we may write for $j \geqslant 1$:

$$E[e^{-\rho s} | x = j] = E[e^{-\rho(\zeta + \tau_1 + \cdots + \tau_x)} | x = j]$$
$$= E[e^{-\rho\zeta} | x = j] E[e^{-\rho(\tau_1 + \cdots + \tau_x)} | x = j]$$
$$= \beta^j(\rho) E[e^{-\rho\zeta} | x = j], \quad \text{Re}\,\rho \geqslant 0. \tag{4.15}$$

It is now trivial to complete the proof.

Remark 4.1. It would be very interesting and useful to obtain the joint distribution of s and z (and x) in the M/G/1/N queue. However, we did not succeed in solving this problem. While the relation between s and z is very simple in the M/G/1 case, it seems to be very complicated here, due to the fact that a customer who meets a full system is not admitted. As pointed out in [3], once the joint distribution of s and z is known, also the joint distribution of the sojourn times—and the cycle time distribution—can be determined in a practically important two-stage cyclic queue (N customers cycle through a closed system of two queues, one queue having a general service time distribution and the other one having a negative exponential service time distribution).

Theorem 4.3.

$$E[x\zeta] = \frac{1}{C} \frac{1}{2\pi i} \int_{D_\omega} \frac{1}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \left[-\beta + \alpha \frac{1 - \beta\left(\frac{1-\omega}{\alpha}\right)}{1-\omega} \right]$$
$$\times \left[\left(1 - \left(\frac{1}{\omega}\right)^{N-1}\right) \frac{\omega}{(1-\omega)^2} + (N-1)\left(\frac{1}{\omega}\right)^{N-1} \frac{1}{1-\omega} \right] d\omega, \quad |\omega| < \mu_0. \tag{4.16}$$

$$E[xs] = E[x\zeta] + \beta E[x^2]. \tag{4.17}$$

Proof. (4.16) follows from Theorem 4.1; for the proof of (4.17) it suffices to refer to (3.11). We note here that $E[x^2]$ can be determined from (4.5), (4.6), (4.9), (4.10); but one may prefer to use a recursive algorithm of Lavenberg [7] to compute $\Pr\{x = j\}$.

Remark 4.2. Taking in (4.14) Re ρ so large that $|s\beta(\rho)| < |\omega|$, Re $(\alpha\rho + \omega - 1) > 0$ for $|\omega| < \mu_0$, one can simplify this relation somewhat (a similar remark holds for e.g. (4.2)):

$$E[s^x e^{-\rho s}]$$

$$= \frac{1}{C}\beta(\rho) + \frac{1}{C}s\beta(\rho)\frac{1}{2\pi i}\int_{D_\omega} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \frac{\beta\left(\frac{1-\omega}{\alpha}\right) - \beta(\rho)}{\alpha\rho + \omega - 1} \frac{1}{s\beta(\rho) - \omega}\, d\omega$$

$$+ \frac{1}{C}s\beta(\rho)\frac{1}{2\pi i}\int_{D_\omega} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \left(\frac{s\beta(\rho)}{\omega}\right)^{N-1} \left[\frac{\beta(\rho) - \omega + \omega - \beta\left(\frac{1-\omega}{\alpha}\right)}{\alpha\rho + \omega - 1}\right] \frac{1}{s\beta(\rho) - \omega}\, d\omega$$

$$= \frac{1}{C}\beta(\rho) - \frac{1}{C}s\beta(\rho)\frac{1 - s\beta(\rho)}{s\beta(\rho) - \beta\left(\frac{1 - s\beta(\rho)}{\alpha}\right)} \frac{\beta\left(\frac{1 - s\beta(\rho)}{\alpha}\right) - \beta(\rho)}{\alpha\rho + s\beta(\rho) - 1}$$

$$- \frac{1}{C}s\beta(\rho)\frac{\alpha\rho}{\alpha\rho + s\beta(\rho) - 1}\left(\frac{s\beta(\rho)}{1 - \alpha\rho}\right)^{N-1}$$

$$+ \frac{1}{C}s\beta(\rho)\frac{1}{2\pi i}\int_{D_\omega} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \left(\frac{s\beta(\rho)}{\omega}\right)^{N-1} \frac{\beta(\rho) - \omega}{\alpha\rho + \omega - 1} \frac{d\omega}{s\beta(\rho) - \omega}, \tag{4.18}$$

Re ρ sufficiently large, $|\omega| < \mu_0$.

Note that for $N \to \infty$, $1/C \to 1 - a$ and (4.18) reduces to the expression for $E[s^x e^{-\rho s}]$ found in Section 3.

Remark 4.3. In general it is not possible to further simplify the contour integrals used above, unless $\beta(\ )$ has a particular form or N is small. As an example consider the case $N = 2$. Then $P(s, \rho)$ becomes (cf. (4.2)):

$$P(s, \rho) = -\frac{1}{C}\frac{1}{2\pi i}\int_{D_\omega} \frac{1-\omega}{\omega - \beta\left(\frac{1-\omega}{\alpha}\right)} \frac{\beta\left(\frac{1-\omega}{\alpha}\right) - \beta(\rho)}{\alpha\rho + \omega - 1} \frac{s}{\omega}\, d\omega$$

$$= \frac{1}{C}\beta^{-1}\left(\frac{1}{\alpha}\right)\frac{\beta\left(\frac{1}{\alpha}\right) - \beta(\rho)}{\alpha\rho - 1}s, \tag{4.19}$$

with Re ρ and D_ω such that $|\omega| < 1 - \alpha\rho$.

Here (cf. (4.3))

$$\frac{1}{C} = \Pr\{x = 0\} = \beta\left(\frac{1}{\alpha}\right), \qquad \Pr\{x = 1\} = 1 - \beta\left(\frac{1}{\alpha}\right). \tag{4.20}$$

Hence

$$E[e^{-\rho\zeta} \mid x = 1] = \frac{\beta\left(\frac{1}{\alpha}\right) - \beta(\rho)}{\alpha\rho - 1}\frac{1}{1 - \beta\left(\frac{1}{\alpha}\right)}. \tag{4.21}$$

From (4.20), (4.21) or from (4.16),

$$E[x\zeta] = \beta - \alpha + \alpha\beta\left(\frac{1}{\alpha}\right), \tag{4.22}$$

$$E[xs] = 2\beta - \alpha + (\alpha - \beta)\beta\left(\frac{1}{\alpha}\right). \tag{4.23}$$

References

[1] S. Asmussen, Equilibrium properties of the M/G/1 queue, *Z. Wahrsch. Verw. Gebiete 58* (1981) 267–281.
[2] R.E. Barlow, A.W. Marshall and F. Proschan, Properties of probability distributions with monotone hazard rate, *Ann. Math. Statist. 34* (1963) 375–389.
[3] O.J. Boxma and P. Donk, On response time and cycle time distributions in a two-stage cyclic queue, *Performance Evaluation 2* (1982) 181–194.
[4] J.W. Cohen, *The Single Server Queue* (North-Holland, Amsterdam, 1981).
[5] D.G. Fakinos, The expected remaining service time in a single server queue, *Operations Res. 30* (1982) 1014–1018.
[6] F.P. Kelly, *Reversibility and Stochastic Networks* (Wiley, New York, 1979).
[7] S.S. Lavenberg, The steady-state queueing time distribution for the M/G/1 finite capacity queue, *Management Sci. 21* (1975) 501–506.
[8] A. Mandelbaum and U. Yechiali, Individual optimization in the M/G/1 queue, *Management Sci.*, to appear.
[9] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications* (Academic Press, New York, 1979).
[10] S.A. Nozaki and S.M. Ross, Approximations in finite-capacity multi-server queues with Poisson arrivals, *J. Appl. Prob. 15* (1978) 826–834.
[11] H.C. Tijms and M.H. van Hoorn, Computational methods for single-server and multi-server queues with random and quasi random input, *Proc. Conference "Applied Probability–Computer Science: The Interface"*, Boca Raton, 1981.
[12] D.M.G. Wishart, Queueing system in which the discipline is 'last-come, first-served', *Operations Res. 8* (1960) 591–599.
[13] D.M.G. Wishart, An application of ergodic theorems in the theory of queues, *Proc. 4th Berkeley Symposium, Vol. II* (1961) 581–592.