



UvA-DARE (Digital Academic Repository)

Conditional ages and residual service times in the M/G/1 queue

Adan, I.J.B.F.; Haviv, M.

[Link to publication](#)

Citation for published version (APA):

Adan, I., & Haviv, M. (2008). Conditional ages and residual service times in the M/G/1 queue. (Eurandom report series; No. 2008-023). Eindhoven: Eurandom.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Conditional ages and residual service times in the M/G/1 queue

Ivo Adan ^{*}and Moshe Haviv ^{†‡}

June 23, 2008

Abstract

In the paper we study the M/G/1 queue, and collect results on the age, residual and length of service, conditional on the number of customers present in the system. Special attention is given to the M/M/1 queue.

1 Introduction

Consider the standard M/G/1 queueing model. The queueing regime can be any non-preemptive, work-conserving and non-anticipating one. For example, it can be first-come first-served. It is well known that the age and the residual of the service length of the customer who is currently receiving service are distributed according to the equilibrium distribution of the service time. However, if additional information is available, such as the number of customers currently present in the system, this is no longer true.

There is, of course, an interest in these conditional distributions. For example, customers who observe the queue length upon arrival and who have to decide whether or not to join queue, need to assess the residual service time of the one in service (on top of how many are in the queue) in order to

^{*}Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands. E-mail: iadan@win.tue.nl

[†]Department of Statistics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. E-mail: haviv@mscc.huji.ac.il.

[‡]This research was supported by The Israel Science Foundation Grant no. 237/02.

be able to estimate their future waiting time; see [10] on the behavior of a ‘smart’ customer, and [1, 5, 8] for non-cooperative games resulting when all customers become smart. Also, there might be an interest in assessing the age of service, when service is given in phases and each phase is performed by a different processor. Hence, knowledge, or estimation, of the age can be utilized in order to set the required processor on time. There is also an interest in assessing the total service time of the one in service (which in fact equals the age plus the residual). This is, for example, the case when customers are paying some amount for service, which is a function of the actual service length.

In this paper we derive the density function and the Laplace-Stieltjes transform (LST) of the conditional age, residual and length of service, given the number of customers in the system. Special attention is given to the case where it is assumed additionally that service times follow an exponential distribution. Some of the reported results are known, but in this paper, all is put in one unified form. In particular, short, straightforward and probabilistic proofs will be given.

The paper is organized as follows. First, in Section 2 we present some preliminaries needed for our derivations. In Section 3 we deal with the conditional age while in Section 4 we do the same regarding the residual service time. Section 5 comes with the conditional total service time. Some concluding remarks and ideas for future research are given in Section 6.

2 Model

We consider the standard M/G/1 queue. Specifically, to a single server queue, there is a Poisson arrival process whose rate is denoted by λ . Service times are independent and their common distribution function is denoted by $G(\cdot)$ with density $g(\cdot)$. The mean service time is denoted by \bar{x} , and LST of $G(\cdot)$ is

$$\tilde{G}(s) = \int_{x=0}^{\infty} e^{-sx} g(x) dx, \quad \text{Re}(s) \geq 0. \quad (1)$$

Service is granted on a first-come first-served basis. As always, for stability, we require that

$$\rho = \lambda \bar{x} < 1,$$

and we assume that the system is in steady-state. Denote by Q the total number of customers in the system (including the one in service) and let

π_n be the steady-state probability that $Q = n$, $n \geq 0$. Recall that these probabilities are also applicable at arrival and departure instants. It is well-known that $\pi_0 = 1 - \rho$. The probability generating function (PGF) of π_n is denoted by

$$P(z) = \mathbb{E}(z^Q) = \sum_{n=0}^{\infty} \pi_n z^n, \quad |z| \leq 1. \quad (2)$$

Also, the famous Pollaczek-Khinchin formula relates the above two transforms (1) and (2) via the following formula:

$$P(z) = (1 - \rho) \frac{(1 - z)\tilde{G}(\lambda(1 - z))}{\tilde{G}(\lambda(1 - z)) - z}. \quad (3)$$

Further, let Q^+ denote the conditional number of customers in the system, given that the server is busy; that is, $Q^+ = Q|Q > 0$, so

$$\mathbb{P}(Q^+ = n) = \frac{\pi_n}{1 - \pi_0} = \frac{\pi_n}{\rho}, \quad n \geq 1.$$

We denote by A and R the random variables which are the age and the residual, respectively, of the service length of the customer who is currently receiving service. Finally, let $L = A + R$ be the total length of service requirement of this customer. We like to remind the reader that A and R are identically distributed (but of course they are not independent). Their density function is $f_A(x) = f_R(x) = (1 - G(x))/\bar{x}$, valid for $x \geq 0$. Hence, their LST equals

$$\tilde{R}(s) = \tilde{A}(s) = \mathbb{E}(e^{-sA}) = \int_{x=0}^{\infty} e^{-sx} f_A(x) dx = \frac{1 - \tilde{G}(s)}{s\bar{x}}.$$

The joint density of A and R is $f_{A,R}(x, y) = g(x + y)/\bar{x}$ and the joint LST is

$$\tilde{J}(s, t) = \mathbb{E}(e^{-sA - tR}) = \frac{\tilde{G}(t) - \tilde{G}(s)}{(s - t)\bar{x}}.$$

The density of function L equals $f_L(\ell) = \ell g(\ell)/\bar{x}$ and its LST is

$$\tilde{L}(s) = -\frac{\tilde{G}^{(1)}(s)}{\bar{x}},$$

where the notation $f^{(n)}(x)$ denotes the n -th derivative of the function $f(x)$.

Initiating with $\pi_0 = 1 - \rho$, the rest of the limit probabilities can be computed via the well-known recursion (see, e.g., [12], p. 178),

$$\pi_{i+1} = \frac{1}{\alpha_0} \left(\pi_i - \pi_0 \alpha_i - \sum_{j=1}^i \pi_j \alpha_{i+1-j} \right), \quad i \geq 0 \quad (4)$$

where α_i stands for the probability that exactly i customers arrive during a single service period. In other words,

$$\alpha_i = \int_{t=0}^{\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t} g(t) dt, \quad i \geq 0.$$

Note that $\alpha_0 = \tilde{G}(\lambda)$, which, coupled with (4) for $i = 0$, leads to

$$\pi_1 = \pi_0 \frac{1 - \alpha_0}{\alpha_0} = (1 - \rho) \frac{1 - \tilde{G}(\lambda)}{\tilde{G}(\lambda)}. \quad (5)$$

An alternative to recursion (4), avoiding subtractions, is (see, e.g., [11])

$$\pi_{i+1} = \frac{1}{\alpha_0} \left(\pi_0 \beta_i + \sum_{j=1}^i \pi_j \beta_{i+1-j} \right), \quad i \geq 0 \quad (6)$$

where $\beta_i = \sum_{j=i+1}^{\infty} \alpha_j$.

In the next section we will start to study the conditional age.

3 Conditional service age

Let $f_{A|Q^+=n}(\cdot)$ denote the conditional density of the age of service given $Q^+ = n$, $n \geq 1$, customers are present in the system (including the one in service). We next give an explicit expression for it for any $n \geq 1$. This result appeared in [2], but below a shorter and more revealing proof is provided. Also, in the case where $n = 1$, expression (9) below agrees with [12], p. 392. The proof is based on the following simple observation; the random variable Q^b denotes the number of customers at the current service commencement, inclusive the one being served.

Lemma 3.1 *The probability distribution of Q^b is equal to*

$$P(Q^b = 1) = \pi_0 + \pi_1, \quad P(Q^b = n) = \pi_n, \quad n \geq 2. \quad (7)$$

Moreover, Q^b is independent of A and of R .

Proof. First recall that the distribution of the number at departure instants in the same as the distribution at a random time. Then, with the exception of $Q^b = 1$, the one who commences service sees Q^b customers (including himself), which is as many as left behind by the previously serviced customer. The case where $Q^b = 1$ is possible when the previously serviced customer leaves behind zero or one customers. The independence of A and of R is obvious, as the service time does not depend on the number in the system. \square

Theorem 3.1 *The conditional density of the age $f_{A|Q^+=n}(\cdot)$, $n \geq 1$, equals*

$$f_{A|Q^+=n}(a) = \frac{\rho}{\pi_n} f_A(a) \left[(1 - \rho) \frac{(\lambda a)^{n-1}}{(n-1)!} + \sum_{i=1}^n \pi_i \frac{(\lambda a)^{n-i}}{(n-i)!} \right] e^{-\lambda a}, \quad a > 0. \quad (8)$$

In particular,

$$f_{A|Q^+=1}(a) = \frac{\rho}{1 - \tilde{G}(\lambda)} f_A(a) e^{-\lambda a} = \lambda \frac{1 - G(a)}{1 - \tilde{G}(\lambda)} e^{-\lambda a}. \quad (9)$$

Proof. For any $n \geq 1$,

$$f_{A|Q^+=n}(a) = \frac{f_A(a) P(Q^+ = n | A = a)}{P(Q^+ = n)}.$$

Note first that $P(Q^+ = n) = \pi_n / \rho$. Hence, it remains to determine $P(Q^+ = n | A = a)$. In order to have n customers in the system at this stage of service, there must have been i customers there upon this service commencement for some i , $1 \leq i \leq n$. Clearly,

$$P(Q^+ = n | A = a) = \sum_{i=1}^n P(Q^b = i) e^{-\lambda a} \frac{(\lambda a)^{n-i}}{(n-i)!}, \quad n \geq 1. \quad (10)$$

Substituting (7) completes the proof. Finally, the special expression for $n = 1$ is based on (5) above. \square

Remark. From (8), we can see that in order to determine $f_{A|Q^+=n}(a)$, one needs to have in hand the values of π_i , $0 \leq i \leq n$. In particular, there is no need to compute in advance the entire stationary distribution of Q in order to compute this conditional density. This is of particular importance, since the vector π_i , $i \geq 0$, can be computed recursively, initiating with $\pi_0 = 1 - \rho$ as can be seen by (4) or (6) above.

Example: M/M/1. In the special case where service is exponentially distributed with rate μ , $\pi_i = (1 - \rho)\rho^i$, $i \geq 0$ and $f_A(a) = \mu e^{-\mu a}$, (8) equals

$$f_{A|Q^+=n}(a) = \mu e^{-(\lambda+\mu)a} \frac{(\mu a)^{n-1}}{(n-1)!} + \lambda e^{-(\lambda+\mu)a} \sum_{i=1}^n \frac{(\mu a)^{n-i}}{(n-i)!}. \quad (11)$$

Also, since $\tilde{G}(s) = \mu/(\mu + s)$, and hence $\tilde{G}(\lambda) = 1/(1 + \rho)$, (9) equals

$$f_{A|Q^+=1}(a) = (\mu + \lambda) e^{-(\lambda+\mu)a}, \quad (12)$$

namely, in the case of an empty queue, the service age of the one in service follows an exponential distribution with parameter $\lambda + \mu$. Note that (12) agrees with [12], p. 392.

Using the identity

$$\int_{a=0}^{\infty} a^{n-i} e^{-(\lambda+s)a} f_A(a) da = (-1)^{n-i} \tilde{A}^{(n-i)}(\lambda + s),$$

it is now straightforward to find the LST of $A|Q^+ = n$, resulting in:

Corollary 3.1 *The LST of $A|Q^+ = n$ equals*

$$\begin{aligned} \mathbb{E}(e^{-sA}|Q^+ = n) &= \int_{a=0}^{\infty} e^{-sa} f_{A|Q^+=n}(a) da \\ &= \frac{\rho}{\pi_n} \left[\pi_0 \frac{(-\lambda)^{n-1}}{(n-1)!} \tilde{A}^{(n-1)}(\lambda + s) + \sum_{i=1}^n \pi_i \frac{(-\lambda)^{n-i}}{(n-i)!} \tilde{A}^{(n-i)}(\lambda + s) \right]. \end{aligned}$$

Example: M/M/1. In this case the LST of the age, given a queue length of n , $n \geq 1$, equals

$$\mathbb{E}(e^{-sA}|Q^+ = n) = \rho^{1-n} \left[\frac{\mu \lambda^{n-1}}{(\mu + \lambda + s)^n} + \sum_{i=1}^n \rho^i \frac{\mu \lambda^{n-i}}{(\mu + \lambda + s)^{n-i+1}} \right].$$

The following result gives the joint transform of A and Q^+ .

Theorem 3.2 For $\operatorname{Re}(s) \geq 0, |z| \leq 1$, the joint transform of A and Q^+ is

$$\mathbb{E}(e^{-sA} z^{Q^+}) = (\pi_0(z-1) + P(z)) \tilde{A}(\lambda(1-z) + s) \quad (13)$$

$$= \frac{(1-\rho)z\tilde{A}(\lambda(1-z) + s)}{1 - \rho\tilde{A}(\lambda(1-z))}. \quad (14)$$

Proof. Recall that Q^b denotes the number of customers present upon the service commencement. The number of Poisson arrivals during the age A is denoted by $N(A)$. Then, $Q^+ = Q^b + N(A)$, where, by Lemma 3.1, Q^b is independent of $N(A)$ and A . Hence,

$$\mathbb{E}(e^{-sA} z^{Q^+}) = \mathbb{E}(e^{-sA} z^{N(A)})\mathbb{E}(z^{Q^b}). \quad (15)$$

By (7) we have

$$\mathbb{E}(z^{Q^b}) = \sum_{i=1}^{\infty} \mathbb{P}(Q^b = i)z^i = \pi_0(z-1) + P(z). \quad (16)$$

Then (13) readily follows by substitution of (16) into (15), together with

$$\begin{aligned} \mathbb{E}(e^{-sA} z^{N(A)}) &= \int_{a=0}^{\infty} e^{-sa} \mathbb{E}(z^{N(a)}) f_A(a) da \\ &= \int_{a=0}^{\infty} e^{-sa} e^{-\lambda a(1-z)} f_A(a) da \\ &= \tilde{A}(s + \lambda(1-z)), \end{aligned}$$

where the second equality is based on the observation that $N(a)$ is Poisson distributed with mean λa , the PGF of which is equal to $e^{-\lambda a(1-z)}$. By use of (3), which can be rewritten as

$$P(z) = \frac{(1-\rho)\tilde{G}(\lambda(1-z))}{1 - \rho\tilde{A}(\lambda(1-z))}, \quad (17)$$

the version (14) readily follows. \square

Example: M/M/1. In the M/M/1 case, formula (14) reduces to

$$\mathbb{E}(e^{-sA} z^{Q^+}) = \frac{(\mu z - \lambda z)(\mu + \lambda - \lambda z)}{(\mu - \lambda z)(\mu + \lambda - \lambda z + s)}.$$

By differentiating (14) with respect to s , multiplying by -1 and setting $s = 0$ yields

$$\begin{aligned} \mathbb{E}(Az^{Q^+}) &= \sum_{n=0}^{\infty} \mathbb{E}(A; Q^+ = n) z^n \\ &= -\frac{(1-\rho)z\tilde{A}^{(1)}(\lambda(1-z))}{1-\rho\tilde{A}(\lambda(1-z))}. \end{aligned} \quad (18)$$

In special cases, transform (18) can be inverted to obtain the conditional expectations:

$$\mathbb{E}(A|Q^+ = n) = \frac{\mathbb{E}(A; Q^+ = n)}{\mathbb{P}(Q^+ = n)}.$$

One example, when the service time follows an exponential distribution, is given next.

Example: M/M/1. For the M/M/1 with arrival rate λ and service rate μ we have

$$\tilde{G}(s) = \tilde{A}(s) = \frac{\mu}{\mu + s}, \quad \pi_n = (1-\rho)\rho^n, \quad n \geq 0,$$

and substitution in (18) gives

$$\begin{aligned} \mathbb{E}(Az^{Q^+}) &= -\frac{(1-\rho)z\frac{-\mu}{(\mu+\lambda(1-z))^2}}{1-\rho\frac{\mu}{\mu+\lambda(1-z)}} \\ &= \frac{(1-\rho)z}{\mu(1-\rho z)(1+\rho(1-z))} \\ &= \frac{1-\rho}{\lambda\rho} \left\{ \frac{1}{1-\rho z} - \frac{1}{1-\rho z/(1+\rho)} \right\}. \end{aligned}$$

Hence,

$$\mathbb{E}(A; Q^+ = n) = (1-\rho)\rho^{n-1} \frac{1}{\lambda} \left(1 - \frac{1}{(1+\rho)^n} \right).$$

Dividing by $P(Q^+ = n) = (1 - \rho)\rho^{n-1}$ for $n \geq 1$, we conclude that

$$E(A|Q^+ = n) = \frac{1}{\lambda} \left(1 - \frac{1}{(1 + \rho)^n} \right), \quad n \geq 1. \quad (19)$$

Hence, the more customers present, the longer the mean service age is. Also, the unconditional mean age equals

$$E(A) = \sum_{n=1}^{\infty} E(A|Q^+ = n)P(Q^+ = n) = \sum_{n=1}^{\infty} \frac{1}{\lambda} \left(1 - \frac{1}{(1 + \rho)^n} \right) (1 - \rho)\rho^{n-1} = \frac{1}{\mu}.$$

Note that the value for $E(A)$ here is not solely due to the memoryless property of service times, but rather due to the time-reversibility of the M/M/1 queue (see also the second proof of Theorem 3.3 below). Specifically, when the orientation of time is reversed, the M/M/1 queue behaves statistically the same (under steady-state conditions). Thus, ages in the original process, correspond to residual service times in the time-reversed process, which, now by the memoryless property, follow the exponential distribution with parameter μ . For more on the concept of time-reversibility, see [7].

Remark. It is clear from (8) that the joint probability-density function $p(n, a)$ for the pair (n, a) , assuming (but not conditioning) $n \geq 1$, equals

$$\begin{aligned} p(n, a) &= P(Q = n)f_{A|Q=n}(a) \\ &= \pi_n f_{A|Q^+=n}(a) \\ &= \rho f_A(a) \left[(1 - \rho) \frac{(\lambda a)^{n-1}}{(n-1)!} + \sum_{i=1}^n \pi_i \frac{(\lambda a)^{n-i}}{(n-i)!} \right] e^{-\lambda a}, \quad a > 0. \end{aligned}$$

An alternative result is given in [12], pp.388–392. Specifically, let

$$G(z, a) = \sum_{n=1}^{\infty} p(n, a)z^n, \quad a > 0.$$

Then, it is shown in [12] that

$$G(z, a) = G(z, 0)e^{-\lambda(1-z)a}(1 - G(a)), \quad a \geq 0$$

where

$$G(z, 0) = (1 - \rho) \frac{\lambda z(1 - z)}{\tilde{G}(\lambda(1 - z)) - z}.$$

In the following subsection we show that, for the special case of the M/M/1 queue, the density of the conditional age can be found directly by employing probabilistic arguments.

3.1 Conditional age for the M/M/1

In this section we consider the special case of exponential service times with rate μ .

Theorem 3.3 *In an M/M/1 queueing system with an arrival rate of λ and a service rate of μ , the conditional age of service $A|Q^+ = n$ is distributed as*

$$A|\{Q^+ = n\} \stackrel{d}{=} \min\{Y, Z(n)\}, \quad n \geq 1, \quad (20)$$

where the random variables Y and $Z(n)$ are independent, Y is exponentially distributed with parameter λ and $Z(n)$ is Erlang- n distributed with scale parameter μ .

Proof 1. The proof is by direct verification. First note that

$$P(\min\{Y, Z(n)\} \geq a) = P(Y \geq a)P(Z(n) \geq a) = e^{-\lambda a} \sum_{i=0}^{n-1} e^{-\mu a} \frac{(\mu a)^i}{i!}.$$

Differentiating and multiplying by -1 yields for the density function of $\min\{Y, Z(n)\}$ that

$$f_{\min\{Y, Z(n)\}}(a) = \lambda e^{-\lambda a} \sum_{i=0}^{n-1} e^{-\mu a} \frac{(\mu a)^i}{i!} + e^{-\lambda a} \mu e^{-\mu a} \frac{(\mu a)^{n-1}}{(n-1)!} \quad (21)$$

which in fact coincides with (11). □

Proof 2. The proof of Theorem 3.3 is technical and hence does not reveal why Equation (20) holds. However, this equation immediately follows from the time-reversibility property which is possessed by the M/M/1 system. Details are given next. In the time-reversed queue-length process, every arrival corresponds to a departure in the original process and vice versa. Moreover, in an M/M/1 queue, the time-reversed queue-length process is statistically

identical to the original process (i.e., the M/M/1 queue process is time-reversible). Now suppose there are n customers in the system, for some $n \geq 1$. Then the age of the one in service is, in the time-reversed process, the residual of the inter-arrival time (which is exponentially distributed with parameter λ). This is true, except when the residual inter-arrival time is greater than the sum of the (residual) service times of the n customers currently in the system. In this case the age is the sum of these n service times. Thus, since exponential random variables are memoryless, we can conclude that the age is the minimum of an exponential random variable with parameter λ and the sum of n independent and exponential random variables, each of which with parameter μ . \square

The representation in Theorem 3.3 leads to the mean conditional age, and this result already appeared above in (19). Two alternative proofs are given below.

Theorem 3.4 *In an M/M/1 queueing system with arrival rate of λ and service rate of μ , the mean conditional age equals*

$$E(A|Q^+ = n) = \frac{1}{\lambda} - \frac{1}{\lambda} \frac{1}{(1 + \rho)^n}, \quad n \geq 1. \quad (22)$$

In particular, $E(A|Q^+ = n)$ is increasing in n .

Proof 2. Denote $E(A|Q^+ = n)$ by a_n , $n \geq 1$. Suppose there are n customers in the system and consider the time-reversed process. Then a_n is the expected time until the first arrival or until the system is empty, whichever happens first. Clearly, the expected time until the first event is $1/(\lambda + \mu)$. The event is an arrival with probability $\lambda/(\lambda + \mu)$, and it is a departure with probability $\mu/(\lambda + \mu)$, in which case $n - 1$ customers remain in the system. Hence, by conditioning on the first event, we get the recursion

$$a_n = \frac{1}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} a_{n-1}, \quad n \geq 1 \quad (23)$$

where $a_0 = 0$. In particular, $a_1 = 1/(\lambda + \mu)$. This is a system of difference equations. A particular solution to this system is the constant $1/\lambda$ and the solution to the homogeneous system is $\mu^n/(\lambda + \mu)^n$, $n \geq 1$. Thus,

$$a_n = \frac{1}{\lambda} + C \frac{\mu^n}{(\lambda + \mu)^n}, \quad n \geq 1,$$

for some constant C . As $a_1 = 1/(\lambda + \mu)$ we conclude that $C = -1/\lambda$. This completes the proof. \square

Remark. Alternatively, it is possible to prove that the expression in (22) solves (23) uniquely by the use of an induction argument initiating with $a_1 = 1/(\lambda + \mu)$.

Proof 3. Theorem 3.4 can also be proved by a straightforward integration,

$$E(A|Q^+ = n) = \int_{a=0}^{\infty} a f_{A|Q^+=n}(a) da$$

where $f_{A|Q^+=n}(a)$ can be read from (21). However, it is even simpler to use

$$E(A|Q^+ = n) = E(\min\{Y, Z(n)\}) = E(Y) - E(Y - Z(n)|Y > Z(n))P(Y > Z(n)),$$

where Y and $Z(n)$ are defined in Theorem 3.3. From the memoryless property of exponential random variables, we conclude that $E(Y - Z(n)|Y > Z(n)) = E(Y) = 1/\lambda$ and further that,

$$P(Y > Z(n)) = \left(\frac{\mu}{\lambda + \mu} \right)^n,$$

from which Theorem 3.4 immediately follows. \square

The following is an immediate corollary of the previous theorem.

Corollary 3.2 *In an M/M/1 queueing system with arrival rate of λ ,*

$$\lim_{n \rightarrow \infty} E(A|Q^+ = n) = \frac{1}{\lambda} .$$

Remark. The above corollary implies that in an M/M/1 queue, no matter how long the queue is, the one in service cannot be blamed for so far holding the server for time longer than (on average) $1/\lambda$. Of course, he will keep holding the server for time which is exponentially distributed with parameter μ . Recall that the unconditional mean age is $1/\mu$ which is of course smaller than $1/\lambda$.

4 Conditional residual service time

Our next step is to find the density of the residual service time, conditioned on $Q^+ = n$, $n \geq 1$. Clearly,

$$f_{R|A=a}(r) = \frac{g(a+r)}{1-G(a)}, \quad a, r \geq 0.$$

Also, given A , the random variables Q^+ and R are independent. Thus, with the aid of Theorem 3.1 we get (as [2]):

Theorem 4.1 *The conditional density of the residual service time $f_{R|Q^+=n}(\cdot)$, $n \geq 1$, equals*

$$\begin{aligned} f_{R|Q^+=n}(r) &= \int_{a=0}^{\infty} f_{A|Q^+=n}(a) f_{R|A=a}(r) da \\ &= \frac{\lambda}{\pi_n} \int_{a=0}^{\infty} \left[\pi_0 \frac{(\lambda a)^{n-1}}{(n-1)!} e^{-\lambda a} + \sum_{i=1}^n \pi_i \frac{(\lambda a)^{n-i}}{(n-i)!} e^{-\lambda a} \right] g(a+r) da, \quad r > 0. \end{aligned} \quad (24)$$

Remark. The remark following Theorem (3.1) applies here too: Only the stationary probabilities π_i , $0 \leq i \leq n$, are needed in order to compute $f_{R|Q^+=n}(r)$.

Remark. For an alternative recursion to compute these conditional densities and their LSTs, which does not call for the prior computation of any stationary probabilities, see [7]. In fact, the recursion in [7] holds also for the case where the arrival rates are queue-length dependent.

Example: M/M/1. In case of exponential service times with parameter μ we have

$$\frac{g(a+r)}{\bar{x}} = g(a)g(r),$$

and then the conditional density of the residual service time given in (24), simplifies to

$$f_{R|Q^+=n}(r) = \frac{\rho}{\pi_n} \left[\pi_0 \alpha_{n-1} + \sum_{i=1}^n \pi_i \alpha_{n-i} \right] g(r),$$

where α_i denotes the probability of i arrivals during a service time. The term between brackets can be recognized as the probability that a departing

customer leaves behind $n - 1$ customers, so it equals π_{n-1} . Hence, using the fact that $\rho\pi_{n-1}/\pi_n = 1$, for the M/M/1 queue we get that

$$f_{R|Q^+=n}(r) = \frac{\rho\pi_{n-1}}{\pi_n}g(r) = g(r),$$

as expected. Indeed, in the M/M/1 model, Q^+ and R are independent. Hence, due to this triviality, we do not exemplify further this section's result for this special case.

The following result gives the joint transform of R and Q^+ . This transform appeared already in [13] (but note that in [13] the transform is with respect to Q and not Q^+ as below).

Theorem 4.2 *For $Re(s) \geq 0, |z| \leq 1$, the joint transform of R and Q^+ is*

$$\mathbb{E}(e^{-sR}z^{Q^+}) = (\pi_0(z-1) + P(z)) \tilde{J}(\lambda(1-z), s) \quad (25)$$

$$= \frac{(1-\rho)z\tilde{J}(\lambda(1-z), s)}{1-\rho\tilde{A}(\lambda(1-z))}. \quad (26)$$

Proof. We have (see the proof of Theorem 3.2),

$$\mathbb{E}(e^{-sR}z^{Q^+}) = \mathbb{E}(e^{-sR}z^{Q^b+N(A)}) = \mathbb{E}(e^{-sR}z^{N(A)})\mathbb{E}(z^{Q^b}).$$

Since $\mathbb{E}(z^{Q^b}) = \pi_0(z-1) + P(z)$ and

$$\begin{aligned} \mathbb{E}(e^{-sR}z^{N(A)}) &= \int_{a=0}^{\infty} \int_{r=0}^{\infty} e^{-sr} \mathbb{E}(z^{N(a)}) f_{A,R}(a, r) da dr \\ &= \int_{a=0}^{\infty} \int_{r=0}^{\infty} e^{-sr} e^{-\lambda a(1-z)} f_{A,R}(a, r) da dr \\ &= \tilde{J}(\lambda(1-z), s), \end{aligned}$$

Equation (25) immediately follows. As before, the final expression utilizes (17). \square

By differentiating (25) with respect to s , multiplying by -1 and setting $s = 0$ yields

$$\begin{aligned} \mathbb{E}(Rz^{Q^+}) &= \sum_{n=0}^{\infty} \mathbb{E}(R; Q^+ = n) z^n \\ &= \frac{\pi_0(z-1) + P(z)}{\rho(1-z)} \left[\bar{x} - \frac{1 - \tilde{G}(\lambda(1-z))}{\lambda(1-z)} \right], \end{aligned}$$

which by substituting the Pollaczek-Khinchin formula (3) reduces to

$$\begin{aligned}
\mathbb{E}(Rz^{Q^+}) &= \frac{(1-\rho)(\rho z - P(z) + \pi_0)}{\lambda\rho(1-z)} \\
&= \frac{1-\rho}{\lambda\rho(1-z)} \left(\sum_{i=1}^{\infty} \pi_i z - \sum_{i=1}^{\infty} \pi_i z^i \right) \\
&= \frac{1-\rho}{\lambda\rho} \sum_{i=1}^{\infty} \pi_i \frac{z - z^i}{1-z} \\
&= \frac{1-\rho}{\lambda\rho} \sum_{i=1}^{\infty} \pi_i \sum_{n=1}^{i-1} z^n \\
&= \frac{1-\rho}{\lambda\rho} \sum_{n=1}^{\infty} z^n \sum_{i=n+1}^{\infty} \pi_i
\end{aligned}$$

Hence, we obtain

$$\mathbb{E}(R; Q^+ = n) = \frac{1-\rho}{\lambda\rho} \sum_{i=n+1}^{\infty} \pi_i, \quad n \geq 1. \quad (27)$$

Dividing the above quantity by $P(Q^+ = n)$ (which equals π_n/ρ), yields the following result, which also appeared in [9, 4]:

Theorem 4.3 *The mean conditional residual service time equals*

$$\mathbb{E}(R|Q^+ = n) = \frac{1-\rho}{\lambda} \frac{1-h_n}{h_n}, \quad n \geq 1, \quad (28)$$

where $h_n = \pi_n/\sum_{i=n}^{\infty} \pi_i$, $n \geq 0$.

Note that here, as in (8), in order to compute $\mathbb{E}(R|Q^+ = n)$, $n \geq 1$, all is required from the stationary probabilities are the first $n+1$ among them, namely, π_i , $0 \leq i \leq n$, which, as denoted above, can be computed recursively starting with $\pi_0 = 1-\rho$. In fact, things simplify even further. Specifically, from (28), and the facts that $\pi_0 = 1-\rho$ and that $\pi_1 = (1-\rho)(1-\tilde{G}(\lambda))/\tilde{G}(\lambda)$ (see (5)), we conclude that

$$\mathbb{E}(R|Q^+ = 1) = \frac{\bar{x}}{1-\tilde{G}(\lambda)} - \frac{1}{\lambda},$$

(an expression which already appeared in [9, 4, 5]). This value can serve as an initial value for a recursive computation for $E(R|Q^+ = n)$. The recursion itself is

$$\begin{aligned}
E(R|Q^+ = n + 1) &= \frac{1 - \rho}{\lambda} \frac{\sum_{i=n+2}^{\infty} \pi_i}{\pi_{n+1}} \\
&= \frac{1 - \rho}{\lambda} \frac{\sum_{i=n+1}^{\infty} \pi_i - \pi_{n+1}}{\pi_{n+1}} \\
&= \frac{\pi_n}{\pi_{n+1}} \frac{1 - \rho}{\lambda} \frac{\sum_{i=n+1}^{\infty} \pi_i}{\pi_n} - \frac{1 - \rho}{\lambda} \\
&= \frac{\pi_n}{\pi_{n+1}} E(R|Q^+ = n) - \frac{1 - \rho}{\lambda}, \quad n \geq 1,
\end{aligned}$$

which is also derived in [7].

Remark. Theorem 4.3 is so clean, suggesting that a much simpler derivation should be feasible. Indeed, it can be established by the application of Little's law as done in [4]. For the sake of completeness, the proof is repeated here. First, Theorem 4.3 can be stated as

$$q_{n+1} = \lambda \pi_n E(R|Q = n) + \lambda q_{n+1} \bar{x}, \quad n \geq 1, \quad (29)$$

where $q_n = \sum_{i=n}^{\infty} \pi_i$, namely the probability that the number of customers in the system is at least n , $n \geq 1$. Obviously, for all $n \geq 1$,

$$E(R|Q = n) = E(R|Q^+ = n).$$

We next prove (29) using Little's law. Specifically, consider position $n + 1$ in the system (or the n -th in the queue) for $n \geq 1$. Note that the server corresponds to position one. The number of customers there is zero or one with probabilities $1 - q_{n+1}$ and q_{n+1} , respectively. Thus, the expected number in this position is q_{n+1} . Assume now that all cross this position (including those who find $n - 1$ or less customers upon arrival and move to their right position in no time). The arrival rate to this position is hence λ . Finally, we look at the expected time spent in this position per customer. Those who arrive and find less than n customers in the system, a fraction of $1 - q_{n-1}$ of the customers, spend zero time there. A fraction π_n of the customers arrive straight there and spend there an expected time of $E(R|Q = n)$. The rest,

a fraction of q_{n+1} , join position $n + 2$ or higher and hence stay in position $n + 1$ a full service period, whose mean is \bar{x} . Thus, by Little's law, we get

$$q_{n+1} = \lambda(\pi_n E(R|Q = n) + q_{n+1}\bar{x}),$$

as promised.

Remark. Denoting by W_q the queueing time (excluding service), we have

$$E(W_q|Q^+ = n) = (n - 1)\bar{x} + \frac{1 - \rho}{\lambda} \frac{1 - h_n}{h_n}, \quad n \geq 1.$$

5 Conditional service length

In this section we derive the distribution of L , the total service time (age plus residual) for the customer currently in service given the number of customers in the system. We like to note that as opposed to the previous two sections, the resulting process $(Q(t), L(t))$, where $Q(t)$ is the number of customers at time t , and where $L(t)$ is the total service requirement for the one being served at time t , is *not* a Markov process.

Theorem 5.1 *The conditional density of the total service time $f_{L|Q^+=n}(\cdot)$, $n \geq 1$, equals*

$$f_{L|Q^+=n}(\ell) = \frac{g(\ell)e^{-\lambda\ell}}{\pi_n} \left[(1 - \rho) \left(1 - \sum_{i=0}^{n-1} \frac{(\lambda\ell)^i}{i!} \right) + \sum_{i=1}^n \pi_i \left(1 - \sum_{j=0}^{n-i} \frac{(\lambda\ell)^j}{j!} \right) \right] \quad (30)$$

Proof. We have

$$\begin{aligned} f_{L|Q^+=n}(\ell) &= \int_{a=0}^{\infty} f_{A|Q^+=n}(a) f_{L|A=a}(\ell) da \\ &= \int_{a=0}^{\ell} f_{A|Q^+=n}(a) f_{L|A=a}(\ell) da \\ &= \int_{a=0}^{\ell} f_{A|Q^+=n}(a) \frac{g(\ell)}{1 - G(a)} da, \end{aligned}$$

which by (8) expands to

$$f_{L|Q^+=n}(\ell) = \frac{\rho g(\ell)}{\pi_n \bar{x}} \int_{a=0}^{\ell} \left[(1 - \rho) \frac{(\lambda a)^{n-1}}{(n-1)!} e^{-\lambda a} + \sum_{i=1}^n \pi_i \frac{(\lambda a)^{n-i}}{(n-i)!} e^{-\lambda a} \right] da.$$

The result now follows from the fact that

$$\int_{a=0}^{\ell} \lambda \frac{(\lambda a)^{n-1}}{(n-1)!} e^{-\lambda a} da = 1 - \sum_{i=0}^{n-1} \frac{(\lambda \ell)^i}{i!} e^{-\lambda \ell}.$$

□

Remark. Note that $f_{L|Q^+=n}(\ell)$, $n \geq 1$, is also function only of π_i , $0 \leq i \leq n$.

Remark. It is clear that

$$\sum_{i=1}^{\infty} \frac{\pi_n}{1 - \pi_0} f_{L|Q^+=n}(\ell) = \frac{\ell g(\ell)}{\bar{x}}, \quad \ell \geq 0.$$

Coupled with Theorem 5.1, this leads to the conclusion that

$$\sum_{n=1}^{\infty} \left[(1 - \rho) \left(1 - \sum_{i=0}^{n-1} \frac{(\lambda \ell)^i}{i!} \right) + \sum_{i=1}^n \pi_i \left(1 - \sum_{j=0}^{n-i} \frac{(\lambda \ell)^j}{j!} \right) \right] e^{-\lambda \ell} = \lambda \ell. \quad (31)$$

Note that the right-hand side is not a function of the service distribution. This, of course, should then be the case regarding the left-hand side. However, from inspecting the left-hand side of (31), this is far from being obvious.

Theorem 5.1 immediately leads to the LST of $L|Q^+ = n$ (cf. Corollary 3.1).

Corollary 5.1 *The LST of $L|\{Q^+ = n\}$ is equal to*

$$\begin{aligned} \mathbb{E}(e^{-sL}|Q^+ = n) &= \frac{1}{\pi_n} \left[\pi_0 \left(\tilde{G}(\lambda + s) - \sum_{i=0}^{n-1} \frac{(-\lambda)^i}{i!} G^{(i)}(\lambda + s) \right) \right. \\ &\quad \left. + \sum_{i=1}^n \pi_i \left(\tilde{G}(\lambda + s) - \sum_{j=0}^{n-i} \frac{(-\lambda)^j}{j!} G^{(j)}(\lambda + s) \right) \right]. \end{aligned}$$

In the same spirit as Theorems 3.2 and 4.2, we can derive the joint transform of L and Q^+ , which is presented in the following theorem.

Theorem 5.2 *For $\operatorname{Re}(s) \geq 0, |z| \leq 1$, the joint transform of L and Q^+ is*

$$\begin{aligned} \mathbb{E}(e^{-sL} z^{Q^+}) &= (\pi_0(z-1) + P(z)) \tilde{J}(s + \lambda(1-z), s) \\ &= \frac{(1-\rho)z \tilde{J}(s + \lambda(1-z), s)}{1 - \rho \tilde{A}(\lambda(1-z))}. \end{aligned}$$

Finally, in order to find $E(Lz^{Q^+})$, one simply needs to sum up $E(Az^{Q^+})$ and $E(Rz^{Q^+})$ as they appear in (18) and in (27), respectively (or one can use the joint transform in Theorem 5.2).

6 Concluding remarks

In this paper we derived for the M/G/1 queueing model, the density functions and the LSTs of the age, residual and length of service for the customer who is currently in service, given the queue length behind him. We also derived the joint transforms of the queue length with any of these three random variables. Special treatment and analysis was given to the M/M/1 case. Some of the reported results were known, but all has been put in one unified form. When different proofs highlighted various probabilistic phenomena we presented them all.

Some interesting questions are still open. For example, what are the distributions of the age and of the residual *queueing time* of a customer who is in line with m customers in front of him (including the one in service) and n behind him? True, the residual queueing time issue can be derived from our above analysis, as it equals $R|Q^+ = n + m + 1$ plus $m - 1$ independent full service times, but this is not the case regarding the age. Another challenging problem is to consider the G/M/1 model. Of course, the residual service time is a trivial task, but this is not the case regarding the age. In particular, distributions at arrival epochs may differ from those at random times. Other possible questions can relate to correlation between random variables. In [3] one can find the correlation between Q and R . A similar question can be asked regarding the correlation between Q and A .

Acknowledgment. The authors would like to thank Yoav Kerner for stimulating and helpful discussions on this subject.

References

- [1] Altman, E. and R. Hassin (2002), “Non-threshold equilibrium for customers joining an M/G/1 queue,” *Tenth International Symposium on Dynamic Games and Applications Workshop*, Saint-Petersburg, Russia.

- [2] Asmussen, S. (1981), “Equilibrium properties of the M/G/1 queue,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. 58, pp. 267–281.
- [3] Boxma, O.J. (1982), “Joint distribution of sojourn time and queue length in the M/G/1 queue with (in)finite capacity,” *European Journal of Operational Research*, Vol. 16, pp. 246–256.
- [4] Fakinos, D. (1982), “The expected remaining service time in a single server queue,” *Operations Research*, Vol. 30, pp. 1014–1118.
- [5] Haviv, M. and Y. Kerner (2007), “On balking from an empty queue,” *Queueing Systems: Theory and Applications*, Vol. 55, pp. 239–249.
- [6] Kelly, F.P. (1979), *Reversibility and Stochastic Networks*, Wiley, New York.
- [7] Kerner, Y. (2008), “The conditional distribution of the residual service time in the $M_n/G/1$ queue,” *Communications in Statistics – Stochastic Models*, (to appear).
- [8] Kerner, Y. (2008), “Equilibrium joining probabilities for an M/G/1 queue,” (submitted for publication).
- [9] Mandelbaum, A. and U. Yechiali (1979), “The conditional residual service time in the M/G/1 queue,” an unpublished manuscript, see <http://www.math.tau.ac.il/uriy/publications.html>.
- [10] Mandelbaum, A. and U. Yechiali (1983) “Optimal entering rules for a customer with wait option at an M/G/1 queue,” *Management Science*, Vol. 29, pp.174–187.
- [11] Ramaswami, V. (1988), “A stable recursion for the steady state vector in Markov chains of M/G/1 type,” *Communications in Statistics – Stochastic Models*, Vol. 4, pp. 183–188
- [12] Ross, S.M. (1996), *Stochastic Processes, 2nd Edition*, Wiley, New York.
- [13] Wishart, D.M.G. (1961), “An application of ergodic theorem in the theory of queues,” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 581–592.