Probability in the Engineering and Informational Sciences, **22**, 2008, 75–80. Printed in the U.S.A.
DOI: 10.1017/S0269964808000053

ON THE LAW OF THE *i*TH WAITING TIME IN A BUSY PERIOD OF $G/M/c$ QUEUES

OPHER BARON
Rotman School of Management
University of Toronto
Toronto, ON, Canada M5S 3E6
E-mail: opher.baron@rotman.utoronto.ca

We use induction to derive the distribution of the waiting time of the ith waiting customer in a busy period for a $G/M/1$ queue with a first come–first serve service. A trivial implication gives the law for the ith waiting time in a busy period for a $G/M/c$ queue. Finally, we use the Lindley recursion to relate our results to the distribution of random walks.

1. INTRODUCTION

Treatments of the $G/M/1$ queue can be found in several books on queuing theory (e.g., Asmussen [2], Cohen [3], and Prabhu [4]). There are alternative methods to explicitly express the steady-state waiting times in a $G/M/1$ queue (e.g., Prabhu [4, p. 109], starting from the transient behavior of the system, or Asmussen [2, p. 228 and 238] based on the Wiener–Hopf factorization identity). The busy period in $G/M/1$ queues was also characterized (e.g., Cohen [3, p. 225]) and, recently, it has been studied by Adan, Boxma, and Perry [1] using the sample path approach.

In this study, we characterize the conditional law of the waiting time of the ith customer within a busy period of a $G/M/1$ queue given that at least i customers were served within this busy period. As we expect, this distribution is a normalized sum of Erlang(μ, j), E_j^{μ}, distributions with $j = 1, \ldots, i-1$. Therefore, our contribution is the explicit expressions for the probability that on arrival, the ith waiting customer in the busy period sees j customers in the system (or, equivalently, $j-1$ customers in the queue).

We also state two straightforward applications for our results. The first is the law for the waiting time of the ith customer in a busy period of a $G/M/c$ queue. The second uses the equivalence of the Lindley recursion, describing the waiting time in a $G/G/1$ queue with a first come–first serve (FCFS) service discipline, to random walks to describe the distribution of random walks condition on that they do not cross a given threshold.

For completeness, we call to mind that the probability density function (pdf) and cumulative distribution function (cdf) of an Erlang(μ, j) random variable, with j phases, each with mean μ^{-1}, are given by

$$f_{E_j^\mu}(x) = e^{-\mu x}\frac{\mu^j x^{j-1}}{(j-1)!}, \quad \forall x \geq 0, \tag{1}$$

$$F_{E_j^\mu}(x) = 1 - e^{-\mu x}\sum_{k=0}^{j-1}\frac{(\mu x)^k}{k!}, \quad \forall x \geq 0, \tag{2}$$

respectively, and the superscript μ emphasizes the mean of each phase.

2. WAITING TIME OF THE *i*TH CUSTOMER WITHIN A BUSY PERIOD

2.1. For the *G/M/1* Queue

Consider a $G/M/1$ queue with interarrival times Z_i for $i \geq 1$ that are independent and identically distributed (i.i.d.) random variables with a cdf $F_Z(z)$ and service requirements S_i for $i \geq 1$ that are i.i.d. and exponentially distributed with mean μ^{-1}.

We let $N + 1$ be the number of customers served during a busy period. We number the arrivals during a busy period such that the customer that initiates the busy period (and does not wait) is the $i = 0$ customer in the busy period. Then the first customer that waits, if such exists, is the $i = 1$ customer and so on until the $i = N$th customer. Note that upon completion of service for the Nth customer, the server is idle. Furthermore, because $E(X) < 0$, busy periods are i.i.d. and $E(N) < \infty$ almost surely.

We let W_i be distributed as the waiting time of the ith customer in a busy cycle on[1]

$$I\{N \geq i\} = 1. \tag{3}$$

Then, the cdf of W_i is defined as

$$F_{W_i}(x) = P(W_i \leq x) = E(I\{0 < W_i \leq x\}), \quad \forall x \geq 0. \tag{4}$$

We let

$$a_i^\mu = \int_0^\infty e^{-\mu t}\frac{(\mu t)^i}{i!}\, dF_Z(t)$$

be the probability that exactly i customers would be served within an interarrival time if there are an infinite number of customers in the queue. Again, the superscript μ emphasizes the mean service time. Moreover, it is clear that $\sum_{i=0}^{\infty} a_i^\mu = 1$.

Using the memoryless of the service time, W_1, the waiting time of the first customer that waits is exponentially distributed with mean μ^{-1}. In the proof of

Theorem 1 we use this fact to express the distribution of W_i, for each $i \geq 1$. Let $\overset{\mathrm{d}}{=}$ denote an equality in distribution; then we have the following theorem.

THEOREM 1: *Let W_i be a random variable with a cdf defined by Eq. (4) for a G/M/1 queue with interarrival times Z with a cdf $F_Z(z)$ and exponential service with mean μ^{-1}. Then*

$$W_i \overset{\mathrm{d}}{=} \sum_{j=1}^{i} P_j^i E_j^{\mu}, \tag{5}$$

where E_j^{μ} is an Erlang(μ, j) *distribution with a pdf and cdf given in Eq. (1), $P_1^1 = 1$ and $P_0^i = 0$ for each $i \geq 1$, and P_j^i for $i \geq 2$ is*

$$P_j^i = \sum_{k=j-1}^{i-1} P_k^{i-1} a_{k-j+1}^{\mu}, \quad \forall j = 1, \ldots, i. \tag{6}$$

PROOF: We argue by induction. For $i = 1$, the distribution of W_1 is exponential. Thus, $P_1^1 = 1$ and $W_1 \overset{\mathrm{d}}{=} E_1^{\mu}$, so the claim holds for $i = 1$. Now, we assume that for $i - 1 \geq 1$,

$$W_{i-1} \overset{\mathrm{d}}{=} \sum_{j=1}^{i-1} P_j^{i-1} E_j^{\mu} \tag{7}$$

and show that the claim holds for i.

We prove the theorem using the following observation. For each $j = 1, \ldots, i$, we have $W_i \overset{\mathrm{d}}{=} E_j^{\mu}$, if two independent events happened. The first is that $W_{i-1} \overset{\mathrm{d}}{=} E_k^{\mu}$ for $k = j - 1, \ldots, i - 1$,[2] and the second event is that during the interarrival time of the ith customer, there were exactly $l = k + 1 - j$ service completions; that is, upon arrival, the $(i - 1)$st customer is the $(k + 1)$st customer in the system (queue + service), and then $k + 1 - j$ service completions take place until the arrival of the ith customer. These two independent events lead to that j customers are seen by the ith arrival. Furthermore, $P(W_{i-1} \overset{\mathrm{d}}{=} E_j^{\mu}) = P_j^{i-1}$ for $k = j - 1, \ldots, i - 1$ and the probability that exactly $k + 1 - j$ services are completed during an interarrival time is a_{k+1-j}^{μ}. Thus, because the ith customer can see j customers in the system only if W_{i-1} saw at least $j - 1$, we have

$$\begin{aligned} P\left(W_i \overset{\mathrm{d}}{=} E_j^{\mu}\right) &= \sum_{k=j-1}^{i-1} P\left(W_{i-1} \overset{\mathrm{d}}{=} E_k^{\mu}\right) a_{k+1-j}^{\mu} \\ &= \sum_{k=j-1}^{i-1} P_k^{i-1} a_{k+1-j}^{\mu}, \end{aligned} \tag{8}$$

where the last equality follows from the induction assumption in Eq. (7). Observing that the ith customer can see 1 to i customers in the system upon arrival completes the proof. ■

We conclude this subsection by establishing that the distribution given for W_i in Theorem 1 is proper. This is equivalent to proving that

$$\sum_{j=1}^{i} P_j^i = \sum_{j=1}^{i} \sum_{k=j-1}^{i-1} P_k^{i-1} a_{k-j+1}^{\mu} = E\big(I\{N \geq i\}\big). \tag{9}$$

To establish Eq. (9), we observe that given $W_{i-1} \stackrel{\mathrm{d}}{=} E_k^{\mu}$, the busy cycle ends if there were $k+1$ or potentially more service completions during the interarrival time of the ith customer. Thus,

$$\begin{aligned} P\big(N \geq i \mid W_{i-1} \stackrel{\mathrm{d}}{=} E_k^{\mu}\big) &= E\Big(I\big\{N \geq i \mid W_{i-1} \stackrel{\mathrm{d}}{=} E_k^{\mu}\big\}\Big) \\ &= \sum_{l=0}^{k} a_l^{\mu}, \end{aligned} \tag{10}$$

and using Eq. (10),

$$\begin{aligned} \sum_{j=1}^{i} \sum_{k=j-1}^{i-1} P_k^{i-1} a_{k-j+1}^{\mu} &= \sum_{k=0}^{i-1} P_k^{i-1} \sum_{j=1}^{k+1} a_{k-j+1}^{\mu} \\ &= \sum_{k=1}^{i-1} P_k^{i-1} \sum_{l=0}^{k} a_l^{\mu} \\ &= \sum_{k=1}^{i-1} P_k^{i-1} P\big(N \geq i | W_{i-1} \stackrel{\mathrm{d}}{=} E_k^{\mu}\big) \\ &= E\,(I\{N \geq i\}), \end{aligned} \tag{11}$$

where the last is from the Total Expectation Theorem. This establishes Eq. (9).

2.2. For *G/M/c* Queues

Consider a $G/M/c$ queue where each server has an exponential service time with mean μ^{-1}, as was investigated by Asmussen [2] and Wolff [6]. The busy period in this queue is the time from an arrival that makes all c servers busy to the first time when only $c-1$ servers are busy. Denoting the waiting time of the ith customer in a busy period of a $G/M/c$ queue by W_i^c, a similar proof to the one of Theorem 1 establishes the following corollary.

Corollary 1: *For a $G/M/c$ queue with interarrival times Z with a cdf $F_Z(z)$ and exponential service with mean μ^{-1}, we have*

$$W_i^c \stackrel{\mathrm{d}}{=} \sum_{j=1}^{i} P_j^i E_j^{c\mu}, \tag{12}$$

where $E_j^{c\mu}$ has an Erlang($c\mu$, j) *distribution with a pdf and cdf given in Eq. (1), $P_1^1 = 1$, and $P_0^i = 0$ for each $i \geq 1$, and P_j^i for $i \geq 2$ is*

$$P_j^i = \sum_{k=j-1}^{i-1} P_k^{i-1} a_{k-j+1}^{c\mu}, \quad \forall j = 1, \ldots, i. \tag{13}$$

2.3. The Busy Period in *G*/*M*/1 Queue and Random Walks

For the $G/M/1$ queue with a FCFS service discipline, let

$$X_i = S_i - Z_{i-1}, \quad \forall i \geq 1, \tag{14}$$

with $Z_0 = 0$. Thus, $\{X_i\}_{i=2}^{\infty}$ is a sequence of i.i.d. random variables. Let X be the generic random variable of this sequence and assume that $E(X) < 0$ (i.e., $E(Z) > 1/\mu$).

Consider the *random walk* $\{V_i\}_{i=1}^{\infty}$ given by $V_n = \sum_{i=1}^{n} X_i$ (with $V_0 \equiv 0$) and observe that because $E(X) < 0$, V tends to $-\infty$ almost surely (e.g., Ross [5]). A *one-sided regulated random walk* $\{Y_i\}_{i=1}^{\infty}$ that is regulated at zero is given by

$$Y_0 = 0 \quad \text{and} \quad Y_{i+1} = \max\{0, Y_i + X_{i+1}\}, \quad \forall i \geq 0. \tag{15}$$

Then from the equivalence of the waiting time in a $G/G/1$ queue with a FCFS service discipline to Eq. (15), known also as the Lindley recursion (e.g., Cohen [3]), and from Theorem 1, Corollary 2 follows.

COROLLARY 2: *For random walks with steps defined by Eq. (14) with S_i that follow and exponential distribution with mean μ^{-1}, the pdfs of $\{Y_i \mid \min_{0<j\leq i}\{Y_j\} > 0\}$* and *$\{V_i \mid \min_{0<j\leq i}\{V_j\} > 0\}$ are identical to those of W_i given in Eq. (5) of Theorem 1.*

Acknowledgment

The author wishes to thank David Perry and an anonymous referee for useful comments on an earlier draft of this paper. The research was supported by an NSERC grant.

Notes

1. We could define $\tilde{W}_i$ such that $P(\tilde{W}_i = 0 \mid N < i) = 1$ and $\tilde{W}_i I\{N \geq i\} = W_i$, but this makes our notation cumbersome.

2. Of course, with $j = 1, k = 1, \ldots, i-1$, so we could define, in general, $k = \max\{j-1, 1\}, \ldots, i-1$. However, we preferred to define $P_0^i = 0$ for each i. Both definitions are equivalent, as is evident in Eq. (8).

References

1. Adan, I., Boxma, O.J., & Perry, D. (2005). The *G*/*M*/1 queue revisited. *Mathematical Methods in Operations Research* 62: 437–452.
2. Asmussen, S. (2003). *Applied probability and queues*, 2nd ed. New York: Springer-Verlag.
3. Cohen, J.W. (1982). *The single server queue*, 2nd ed. Amsterdam: North-Holland.

4. Prabhu, N.U. (1965). *Queues and inventories*. New York: Wiley.
5. Ross, S.M. (1983). *Stochastic processes*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.
6. Wolff, R.W. (1988). *Stochastic modeling and the theory of queues*. Prentice-Hall International Series in Industrial and Systems Engineering. Englewood Cliffs, NJ: Prentice-Hall.