Stationary remaining service time conditional on queue length

Karl Sigman* Uri Yechiali †

October 7, 2006

Abstract

In Mandelbaum and Yechiali (1979) a simple formula is derived for the expected stationary remaining service time in a FIFO M/G/1 queue, conditional on the number of customers in the system being equal to j, $j \geq 1$. Fakinos (1982) derived a similar formula using an alternative method. Here we give a short proof of the formula using *rate conservation law (RCL)*, and generalize to handle higher moments which better illustrates the advantages of using RCL.

Keywords M/G/1; CONDITIONAL RESIDUAL SERVICE TIME; RATE CONSERVATION LAW

AMS Subject Classification:

*Department of IEOR, Columbia University, 500 West 120th St., New York, NY 10027, USA

†Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel

1 Introduction

For a stable FIFO M/G/1 queue in (time) stationarity, let $R_j = (S_r|L = j)$ denote the remaining service time of the customer in service conditional that the number of customers in the system (L) equals $j,\ j \geq 1$. Let λ denote the arrival rate, $E(S) = 1/\mu$ denote mean service time and let $\rho \stackrel{\text{def}}{=} \lambda/\mu < 1$.

Mandelbaum and Yechiali [1979] proved that

$$E(R_j) = \frac{1-\rho}{\lambda} \frac{P(L > j)}{P(L = j)}. \tag{1.1}$$

Fakinos [1982] derived similar relationships using different methods. Recently, in Ross[2006], this result is used for purposes of obtaining bounds on stationary queue length.

A basic application is that an arriving customer who finds j customers in the system upon arrival, experiences a delay as precisely the sum $R_j + S_1 + \cdots + S_{j-1}$, where the S_i are iid regular (unbiased) service times of the customers waiting in line. Thus the conditional expected delay can be computed if $E(R_j)$ can be computed. Examples include call centers where an estimate of expected delay needs to be given to an arriving customer; further examples involve admission control (see for example Mandelbaum and Yechiali [1983]).

Here, in the present paper, we give a short proof of (1.1) using *rate conservation law (RCL)*, $E(X') = \lambda E(J)$; see Section 5.5 (Theorems 5.5 and 5.6) in Sigman[1995]. ($E(X')$ denotes time average right derivative of a process $\{X(t)\}$, λ is the rate at which jumps occur for the process, and $-E(J)$ is the average jump size, all of which can be viewed as sample-path averages.) We then generalize to derive higher moments.

2 Proof of (1.1)

Proof : Let $S_r(t)$ denote remaining service time of the served customer at time t, $L(t)$ the number of customers in the system at time t, and $\pi_j = P(L = j)$; recall from PASTA that π_j is both the proportion of time there are j customers in the system and the proportion of arrivals who find j customers in the system. We will proceed by induction: Suppose that the result holds for some $j \geq 1$. We will show that the result then holds for $j+1$ too. To this end let $X(t) = S_r(t)I\{L(t) = j+1\}$ and observe that $X'(t) = -I\{L(t) = j+1\}$. There are three sources of jumps for the process $X(t)$,

1. Arrivals who find $j+1$ customers in the system. The rate of such jumps is given by $\lambda_1 = \lambda\pi_{j+1}$ and a jump size is of the form $-J(1) = X(0+) - X(0-) = 0 - R_{j+1} = -R_{j+1}$ (by PASTA).

2. Arrivals who find j customers in the system. The rate of such jumps is given by $\lambda_2 = \lambda\pi_j$ and a jump size is of the form $-J(2) = X(0+) - X(0-) = R_j - 0 = R_j$ (again by PASTA).

3. Departures who leave $j+1$ customers behind. The rate of such jumps is given by $\lambda_3 = \lambda\pi_{j+1}$ because (via basic sample-path principles[1]) the proportion of departures who leave j customers behind is π_j and the long-run departure rate is λ. A jump size is of the form $-J(3) = X(0+) - X(0-) = S$, a typical service time with mean $E(S) = 1/\mu$.

RCL thus takes the form $E(X') = \lambda_1 E(J(1)) + \lambda_2 E(J(2)) + \lambda_3 E(J(3))$ and becomes

$$\pi_{j+1} = -\lambda\pi_{j+1}E(R_{j+1}) + \lambda\pi_j E(R_j) + \lambda\pi_{j+1}E(S). \tag{2.2}$$

[1] A function that has jumps of only magnitude 1 (such as, but not restricted to, the sample paths of a birth and death process) must satisfy "the long-run rate at which the function jumps from state j to $j+1$ equals the long-run rate at which it jumps from $j+1$ to j. Formally this is sometimes referred to as Burke's theorem.

Plugging in the induction hypothesis (1.1) for $E(R_j)$ and solving for $E(R_{j+1})$ then yields

$$\begin{aligned} E(R_{j+1}) &= \frac{1}{\lambda\pi_{j+1}}\,[(1-\rho)P(L>j)-(1-\rho)\pi_{j+1}] && (2.3)\\ &= \frac{1}{\lambda\pi_{j+1}}\,[(1-\rho)P(L>j+1)] && (2.4)\\ &= \frac{1-\rho}{\lambda}\,\frac{P(L>j+1)}{P(L=j+1)}; && (2.5) \end{aligned}$$

we arrive at (1.1) for $j+1$. Thus it now suffices to prove (1.1) when $j=1$. Repeating the above RCL analysis with $X(t)=S_r(t)I\{L(t)=1\}$, the only modification needed is that the type 2 arrivals (rate $\lambda_2=\lambda\pi_0$) find the system empty and hence enter service immediately; $-J(2)=X(0+)-X(0-)=S$. Thus RCL yields

$$\begin{aligned} \pi_1 &= -\lambda\pi_1 E(R_1)+\lambda\pi_0 E(S)+\lambda\pi_1 E(S)\\ &= -\lambda\pi_1 E(R_1)+\rho(\pi_0+\pi_1). \end{aligned}$$

Solving for $E(R_1)$ while using the fact that $\pi_0=1-\rho$ and $P(L>0)=\rho$ then yields

$$E(R_1)=\frac{1-\rho}{\lambda}\,\frac{P(L>1)}{P(L=1)};$$

the proof is now complete. ■

3 Second moments and beyond

Using the first moment formula (1.1) we can "bootstrap" it via RCL to obtain second moments, $E(R_j^2)$ and then use second moments to obtain third moments and so on. It is this aspect of the RCL method that better illustrates its efficiency. We briefly illustrate the method by obtaining $E(R_1^2)$ and then, more generally, $E(R_1^n)$, and then $E(R_2^n)$. Finally we show how to use Laplace transforms to obtain such recursions.

Let $X(t) = S_r^2(t)I\{L(t) = 1\}$ and observe that $X'(t) = -2S_r(t)I\{L(t) = 1\}$. The same jump sources as for deriving (1.1) apply here. Moreover

$$E(R_1) = \frac{E(S_r I\{L = 1\})}{\pi_1}.$$

RCL thus yields

$$E(R_1) = -\frac{\lambda}{2}E(R_1^2) + \frac{\lambda\pi_0}{2\pi_1}E(S^2) + \frac{\lambda}{2}E(S^2),$$

and we can solve for $E(R_1^2)$ in terms of $E(R_1)$ derived earlier:

$$E(R_1^2) = (1 + \frac{\pi_0}{\pi_1})E(S^2) - \frac{2}{\lambda}E(R_1). \tag{3.6}$$

Continuing recursively for higher moments, $E(R_1^n)$, we use $X(t) = S_r^n(t)I\{L(t) = 1\}$ which then yields

$$E(R_1^n) = (1 + \frac{\pi_0}{\pi_1})E(S^n) - \frac{n}{\lambda}E(R_1^{n-1}),\ n \geq 2. \tag{3.7}$$

Now that we have all first moments $E(R_j)$ for any $j \geq 1$ and all moments for $j = 1$, we can obtain all moments for $j = 2$ by using $X(t) = S_r^n(t)I\{L(t) = 2\}$:

$$E(R_2^n) = \frac{\pi_{j-1}}{\pi_j}E(R_1^n) + E(S^n) - \frac{n}{\lambda}E(R_2^{n-1}),\ n \geq 2. \tag{3.8}$$

Interestingly, all these moment recursions can be obtained by first deriving expressions for the Laplace transform. For a non-negative random variable X, let $\mathcal{L}_X(s) = E(e^{-sX}),\ s \geq 0$, denote the Laplace transform of X. Further, let X_e denote a random variable with the equilibrium distribution (stationary excess) of X, with density $P(X > x)/E(X),\ x \geq 0$. Then, for example, using RCL on $X(t) = e^{-sS_r(t)}\{L(t) = 1\}$ leads to

$$\mathcal{L}_{R_1}(s) + \lambda E(R_1)\mathcal{L}_{R_{1,e}}(s) = \rho(1 + \frac{\pi_0}{\pi_1})\mathcal{L}_{S_e}(s). \tag{3.9}$$

(Here, $R_{1,e} = (R_1)_e$.) Setting $s = 0$ yields $E(R_1)$ and then taking first derivatives with respect to s and setting $s = 0$ yields $E(R_1^2)$, and so on.

References

[1] D. Fakinos (1982). The expected remaining service time in a single-server queue. *Operations Research*, **30**, 1014-1018.

[2] A. Mandelbaum and U. Yechiali (1983). Optimal entering rules for a customer with wait option at an M/G/1 Queue. *Management Science* **29**, 174-187.

[3] A. Mandelbaum and U. Yechiali (1979). The conditional residual service time in the M/G/1 queue. (http://www.math.tau.ac.il/ ∼uriy/publications, (No. 30a).

[4] S. Ross (2006). Bounding the stationary distribution of the M/G/1 queue size. (manuscript)

[5] K. Sigman (1995). *Stationary Marked Point Processes: An Intuitive Approach*. Chapman and Hall, New York.