Two Things About the M/G/1 Queue

The M/G/1 queue is a generalization of the M/M/1 queue where the service times are iid with some distribution with pdf f and cdf F. Keeping with our previous queueing terminology/notation, we will use μ to denote the rate of service. (As opposed to the mean service time. Note that if customers are being served at a rate of μ per unit time then the average (mean) service time is $1/\mu$ units of time per customer. So, $1/\mu = \int_0^\infty x\, f(x)\, dx$.)

$\boxed{1.}$ $\boldsymbol{\pi_0}$

We know that, for the M/M/1 queue, the long-run proportion of time that the queue is empty is $\pi_0 = 1 - \lambda/\mu$. It turns out that this is also the case for the post-departure embedded M/G/1 queue. Here, λ denotes the arrival rate and μ denotes the service rate. You could show this result using the stationary equations, but it is very messy and involves Laplace transforms. Alternatively, you could show it using something called "renewal theory". This is the subject of Chapter 3 of your textbook (Durrett). I am going to find π_0 for the embedded M/G/1 queue here using this approach without actually going to the trouble of introducing renewal theory. (ie: Check out Chapter 3 if you want more details!)

A "busy period" in queueing theory refers to a continguous segment of time in which the server is working.

An "idle period" refers to a continguous segment of time in which the server is not working.

All single server queueing models can be viewed as a sequence of busy periods followed by idle periods:

$$B_n + I_n$$

where B_n is the length of the nth busy period and I_n is the length of time the server is then idle before the next busy period starts.

Suppose that the system is in equilibrium. Let B and I denote typical busy and idle times for the server. The proportion of time the system is empty is

$$\pi_0 == \frac{\mathsf{E}[I]}{\mathsf{E}[B] + \mathsf{E}[I]}.$$

That first equality is a key result in renewal theory. I hope that, in the absence of renewal theory, but it should make some intuitive sense to you even without renewal theory. The long-run proportion of idle time in the system is equivalent to the fraction of idle time in any one of these "busy-then-idle" periods.

We know that, due to the exponential rate λ interarrivals,

$$\mathsf{E}[I] = \frac{1}{\lambda}.$$

Now a busy period consists of a first service time S of the customer who first breaks an idle period, plus busy periods generated by all customers who arrive during this service time. Let N be this number of arrivals.

Okay, the server is idle and a customer arrives. If no one arrives during the service of this customer, the busy time of the server is simply this one service time.

$$\mathsf{E}[B|S_1 = s, N = 0] = s$$

If one customer arrives during the service time of this customer, then that arriving customer generates his/her own busy period (made up of his/her service time and those of the arrivals within that service time).

$$\mathsf{E}[B|S = s, N = 1] = s + \mathsf{E}[B]$$

$$\mathsf{E}[B|S = s, N = 1] = s + 2\mathsf{E}[B]$$

etc...

$$\vdots$$

$$\mathsf{E}[B|S = s, N = n] = s + n \cdot \mathsf{E}[B_1]$$

So,

$$\begin{aligned} \mathsf{E}[B|S = s] &= \textstyle\sum_{n=0}^{\infty} \mathsf{E}[B|S = s, N = n] \cdot P(N = n|S = s) \\ &= \textstyle\sum_{n=0}^{\infty} (s + n \cdot \mathsf{E}[B]) \cdot \frac{e^{-\lambda s}(\lambda s)^n}{n!} \\ &= s + \lambda s \mathsf{E}[B] \end{aligned}$$

Hence, we have

$$\begin{aligned}
\mathsf{E}[B] &= \textstyle\int_0^\infty \mathsf{E}[B|S=s]\, f(s)\, ds \\
&= \textstyle\int_0^\infty (s + \lambda s \mathsf{E}[B])\ f(s)\, ds \\
&= (1+\lambda \mathsf{E}[B])\ \textstyle\int_0^\infty s\, f(s)\, ds \\
&= (1+\lambda \mathsf{E}[B]) \cdot \mathsf{E}[S] \\
&= (1+\lambda \mathsf{E}[B]) \cdot \tfrac{1}{\mu}
\end{aligned}$$

Solving for $\mathsf{E}[B]$, we get

$$\mathsf{E}[B] = \frac{1/\mu}{1-\lambda/\mu} = \frac{1}{\mu-\lambda}.$$

Therefore

$$\pi_0 = \frac{\mathsf{E}[I]}{\mathsf{E}[B]+\mathsf{E}[I]} = \frac{1/\lambda}{\frac{1}{\mu-\lambda}+1/\lambda} = 1 - \frac{\lambda}{\mu}$$

Cool no? Yes!

$\boxed{2.}$ Mean Queue Length in Equilibrium

Let $X(t)$ be the number of customers in the queue at time t. We want to find

$$L = \lim_{t\to\infty} \mathsf{E}[X(t)].$$

Let $\{X_n\}$ be the embedded "post-departure" chain that we talked about in class. (ie: X_n is the number of customers that nth departing customer sees when he/she looks back as he/she is leaving) We saw that $\{X_n\}$ is a Markov chain.

In class we talked briefly about the PASTA property for systems with Poisson arrivals. It says that, for such systems, the proportion of time that an arrival sees n customers in the system is the same as the long-run proportion of time that there are n customers in the continuous-time system. Since an arriving customer causes a "jump" (of size 1) in the state of the system, we often call these observations of arriving customers a "pre-jump chain". There is an analogue to the PASTA property for the "post-departure" chain.

By the PASTA property, we have that

$$L = \lim_{t \to \infty} \mathsf{E}[X(t)] = \lim_{n \to \infty} \mathsf{E}[X_n].$$

Let X be a value of the $\{X_n\}$ chain in equilibrium. (So, you can think of X as X_∞.)

Let X' be the next value of the chain in equilibrium. (So, you can think of X' as $X_{\infty+1}$.)

Then X and X' have the same (stationary) distribution and, in particular,

$$L = \mathsf{E}[X] = \mathsf{E}[X'].$$

Note that we can write

$$X' = X + N - I_{\{X>0\}} \tag{1}$$

where N is the number of arrivals during a service period and $I_{\{X>0\}}$ is the indicator function that takes the value 1 when $X > 0$ and is 0 otherwise.

Taking expectations, we have

$$\mathsf{E}[X'] = \mathsf{E}[X] + \mathsf{E}[N] - \mathsf{E}[I_{\{X>0\}}]$$

Since $\mathsf{E}[X]$ and $\mathsf{E}[X']$ are both L, we can cancel those expectations from both sides of the equation and we get

$$\mathsf{E}[N] = \mathsf{E}[I_{\{X>0\}}]$$

Now

$$\mathsf{E}[N] = \mathsf{E}[I_{\{X>0\}}] = P(X > 0) = 1 - P(X = 0) = 1 - \pi_0 = 1 - (1 - \lambda/\mu) = \lambda/\mu.$$

We now have the expected value of one of the terms in equation (1). However, since the L's canceled out when we took the expected value on both sides of (1), we can't really use this to find L.

So, let's consider squaring both sides of (1):

$$\begin{aligned}
(X')^2 &= (X + N - I_{\{X>0\}})^2 \\
&= X^2 + 2XN - 2XI_{\{X>0\}} - 2NI_{\{X>0\}} + N^2 + I_{\{X>0\}}
\end{aligned}$$

(Here I have used the fact that $I^2_{\{X>0\}} = I_{\{X>0\}}$.)

Note that $XI_{\{X>0\}} = X$ since the indicator is 0 when $X = 0$ which means that the X out front would be zero anyway!

So, we have

$$(X')^2 = X^2 + N^2 + I_{\{X>0\}} - 2X + 2N(X - I_{\{X>0\}}) \tag{2}$$

Will will take expectations all the way through equation (2). Since X and X' have the same distribution, $\mathsf{E}[X^2] = \mathsf{E}[(X')^2]$, so these terms will cancel. Using again the fact that $\mathsf{E}[X] = L$, we have

$$0 = \mathsf{E}[N^2] + \frac{\lambda}{\mu} - 2L + 2\mathsf{E}[N(X - I_{\{X>0\}})]$$

By independence of X and N (Why?), we have

$$\mathsf{E}[N(X - I_{\{X>0\}})] = \mathsf{E}[N] \cdot \mathsf{E}[X - I_{\{X>0\}}] = \frac{\lambda}{\mu} \cdot \left(L - \frac{\lambda}{\mu}\right)$$

so we have

$$0 = \mathsf{E}[N^2] + \frac{\lambda}{\mu} - 2L + 2\frac{\lambda}{\mu} \cdot \left(L - \frac{\lambda}{\mu}\right).$$

Solving for L gives us that

$$L = \frac{\mathsf{E}[N^2] + \frac{\lambda}{\mu} - 2\left(\frac{\lambda}{\mu}\right)^2}{2\left(1 - \frac{\lambda}{\mu}\right)} \tag{3}$$

It remains only for us to find $\mathsf{E}[N^2]$ where N is the number of arrivals during a service time S with pdf f. Since $N|S = s \sim Poisson(\lambda s)$, we have

$$\begin{aligned}
\mathsf{E}[N^2] &= \textstyle\int_0^\infty \mathsf{E}[N^2|S=s]\, f(s)\, ds \\
&= \textstyle\int_0^\infty Var[N|S=s]\, f(s)\, ds - \int_0^\infty \left(\mathsf{E}[N|S=s]\right)^2\, f(s)\, ds \\
&= \lambda \textstyle\int_0^\infty s\, f(s)\, ds + \lambda^2 \int_0^\infty s^2\, f(s)\, ds \\
&= \lambda \mathsf{E}[S] + \lambda^2 \mathsf{E}[S^2] \\
&= \lambda \mathsf{E}[S] + \lambda^2 \left(Var[S] + (\mathsf{E}[S])^2\right) \\
&= \lambda \cdot \tfrac{1}{\mu} + \lambda^2 \left(\sigma^2 + \left(\tfrac{1}{\mu}\right)^2\right)
\end{aligned}$$

where σ^2 is the variance of the service time distribution.

Plugging this back into (3) and simplifying gives us

$$L = \frac{2\frac{\lambda}{\mu} + \lambda^2\sigma^2 - \left(\frac{\lambda}{\mu}\right)^2}{2\left(1 - \frac{\lambda}{\mu}\right)}.$$

More cool!