A NOVEL APPROACH FOR CALCULATING THE QUEUE LENGTH DISTRIBUTION IN THE M/G/1 QUEUE

Matthew Roughan
Department of Applied Mathematics,
University of Adelaide

Abstract: The analytic solution to the M/G/1 queue has long been known. The most common derivations give probability generating functions for the queue length. This approach can have difficulties in the numerical accuracy of the probabilities of large numbers of customers. In this paper the generating function solution to a modified M/G/1 queue is used in conjunction with Little's law to derive these probabilities.

1 Introduction

The solution to the stationary M/G/1 queue has long been known (Cooper, 1972). Many approaches give the solution in generating function form. From this it is not easy to obtain analytic probabilities for the number of customers in the system. These are desirable for a number of reasons, numerical accuracy of calculations being one of them. When modifications are made to the system, such as incorporating a warmup time ,it may become impossible to obtain such probabilities by normal techniques. We will use a modified M/G/1 queue to obtain explicit probabilities for the normal queue. This paper is arranged as follows. In the next section the definitions of the systems we will examine are presented along with the solutions which are derived elsewhere. Following this we use these solutions along with Little's law to derive the probabilities in the M/G/1 queue.

2 Definitions

By the M/G/1 queue, we mean the single server system with a potentially infinite queue to which arrivals come in a Poisson stream with rate λ and service times are independent, identically distributed random variables with probability distribution function $A(\cdot)$ and mean $1/\mu$. Customers who find the server unoccupied seize it immediately and hold it for their service-time. Customers who find the server busy wait in the queue until they receive service. The order of service, or the queue discipline is irrelevant so long as it is noted that it is non-preemptive. In order to obtain the solution the queue is observed immediately after services. PASTA (Wolff, 1989) and the fact that in equilibrium arrivals to the queue see the same distribution that departures leave (Cooper, 1972) mean that the equilibrium distribution of the embedded process, is the same as the equilibrium distribution for the system. The probability generating function for the equilibrium behaviour of the system is then

$$g(z) = (1-\rho)\frac{a(z)(1-z)}{a(z)-z},$$

where $\rho = \frac{\lambda}{\mu}$ and $a(z)$ is the probability generating function for the number of arrivals during one service time which is given by

$$a(z) = \sum_{i=0}^{\infty} a_i z^i,$$

where a_i is the probability of i arrivals during one service and $a(z)$ is given in terms of the Laplace-Stieltjes transform of $A(\cdot)$ by

$$a(z) = A^*(\lambda(1-z)).$$

We will also use the solution to a M/G/1 queue which is modified as follows. The arrivals are again Poisson with rate λ but the service-times are random variables which may take one of two possible probability distribution functions $A(\cdot)$ or $B(\cdot)$. If there are less than $k+1$ customers in the system immediately before a service the service time takes distribution $A(\cdot)$ while if there are more than k customers in the system immediately before the service begins the service-time takes distribution $B(\cdot)$. Using a martingale technique devised by Baccelli and Makowski (1985,1989) the probability generating function for this process in equilibrium is given in Roughan (1993) as

$$E\left[z^X\right] \quad = \quad \frac{1}{m}\left[\frac{b(z)(1-z) + (b(z)-a(z))f_k(z)}{b(z)-z}\right],$$

where

$$\begin{aligned}
a(z) &= A^*(\lambda(1-z)), \\
b(z) &= B^*(\lambda(1-z)), \\
m &= \frac{1 + (\rho_a - \rho_b) f_k(1)}{1-\rho_b}, \\
f_k(z) &= \mathbf{v}(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{z}^t,
\end{aligned}$$

where $\mathbf{z} = (z, z^2, \ldots, z^k)$ and

$$\mathbf{P}_k = \begin{pmatrix}
a_1 & a_2 & a_3 & \cdots & a_{k-1} & a_k \\
a_0 & a_1 & a_2 & \cdots & a_{k-2} & a_{k-1} \\
0 & a_0 & a_1 & \cdots & a_{k-3} & a_{k-2} \\
 & & & \vdots & & \\
0 & 0 & 0 & \cdots & a_0 & a_1
\end{pmatrix}$$

and $\mathbf{v} = (\mathbf{e}_1 + p\mathbf{e}_k)$ where

$$p = \frac{(1 - a_0\mathbf{e}_1(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{e}_k)(1 - a_0\mathbf{e}_k(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{e}_k)}{a_0\mathbf{e}_k(\mathbf{I} - \mathbf{P}_k)^{-1}\mathbf{e}_k},$$

and $\rho_a = \lambda/\mu_a$, $\rho_b = \lambda/\mu_b$ are the mean number of arrivals during a service with distributions $A(\cdot)$ and $B(\cdot)$ respectively.

3 Results

Little's law (1961) states

$$L = \lambda W, \tag{1}$$

where L is the mean number of customers in the system, λ is the arrival rate to the system and W is the mean time spent by a customer in the system. If we apply this to the server alone we can see that L is the probability that there is a customer in the system and W is the mean service time so

$$\begin{aligned} L &= p\{X \neq 0\} = 1 - \frac{1}{m}, \\ W &= \frac{1-p_k}{\mu_a} + \frac{p_k}{\mu_b}. \end{aligned}$$

where p_k is the probability of more than k customers being in the system. Thus we get

$$\begin{aligned} L &= \frac{\rho_b + (\rho_a - \rho_b) f_k(1)}{1 + (\rho_a - \rho_b) f_k(1)}, \\ \lambda W &= p_k(\rho_b - \rho_a) + \rho_b. \end{aligned}$$

Substituting in (1) we get an equation for p_k

$$p_k = \left[\frac{\rho_b + (\rho_a - \rho_b) f_k(1)}{1 + (\rho_a - \rho_b) f_k(1)} - \rho_a \right] \frac{1}{\rho_b - \rho_a}.$$

If we consider ρ_a to be a constant and p_k to be a function of ρ_b we can use L'Hopitals rule to take the limit as $\rho_b \to \rho_a$ to get

$$\begin{aligned} \lim_{\rho_b \to \rho_a} p_k(\rho_b) &= \frac{d}{d\rho_b} \left[\frac{\rho_b + (\rho_a - \rho_b) f_k(1)}{1 + (\rho_a - \rho_b) f_k(1)} - \rho_a \right]_{\rho_b = \rho_a} \\ &= \left[\frac{(1 - f_k(1))(1 + (\rho_a - \rho_b) f_k(1)) + f_k(1)(\rho_b + (\rho_a - \rho_b) f_k(1))}{(1 + (\rho_a - \rho_b) f_k(1))^2} \right]_{\rho_b = \rho_a} \\ &= 1 - f_k(1) + \rho_a f_k(1). \end{aligned}$$

4 Conclusion

This gives us the probability that there are more than k customers in the queue. This works for all $k > 1$ and the case with $k = 0$ is trivial and so we have the required result. This method could be extended to cover far more complex situations. It is not intended to be the be all and end all of such problems. It is merely a demonstration that generating functions do not necessarily obscure the probabilities involved in a problem.

5 Acknowledgements

This work has been funded by an APRA and the Teletraffic Research Center at Adelaide University. I would like to thank my supervisor As. Prof. C.E.M. Pearce for many helpful comments and suggestions.

6 References

F.Baccelli and A.M.Makowski, Direct Martingale Arguments For Stability: the M/GI/1 case, *Systems Control Letters*, (1985), **6**, 181-186.

F.Baccelli and A.M.Makowski, Dynamic, Transient and Stationary Behaviour of the M/GI/1 Queue via Martingales, *Annals of Probability*, (1989), **17**, 4, 1691-1699.

R.B. Cooper, *Introduction to Queueing Theory*, The Macmillian Company, (1972)

J.D.C. Little, A proof of the queueing formula: $L = \lambda W$, *Operations Research*, (1961), **9**, 3, 383-87.

M. Roughan, A martingale technique for the solution of a modified M/G/1 queue, *unpublished.*

R.W.Wolff, *Stochastic Modelling and the theory of Queues*, Prentice Hall International, (1989).