SERIE RESEARCH MEMORANDA

AN ELEMENTARY PROOF OF A BASIC RESULT
FOR THE GI/G/1 QUEUE

J.C.W. van Ommeren

R.D. Nobel

Researchmemorandum 1988-20

March 1988

VRIJE UNIVERSITEIT
Faculteit der Economische Wetenschappen en Econometrie
AMSTERDAM

An elementary proof of a basic result for the GI/G/1 queue.

By
J. C. W. van Ommeren and R. D. Nobel,
Operations Research Group,
Dept. of Econometrics,
Vrije Universiteit,
Postbus 7161,
1007 MC Amsterdam.

Abstract.

The purpose of this note is to give an elementary proof of the result that the complementary waiting-time distribution function in the GI/G/1 queue is the sum of two exponential functions when the service time has a Coxian-2 distribution. Apart from its didactic interest the result is quite useful in the analysis of more complex queueing systems.

1. Introduction.

An important model is the GI/G/1 queue in which the service time has a Coxian-2 distribution. That is, the service time S can be represented as

$$S = \begin{cases} S_1 & \text{with probability } 1-b \\ S_1 + S_2 & \text{with probability } b \end{cases} \tag{1}$$

where S_1 and S_2 are independently exponentially distributed random variables with respective means $1/\mu_1$ and $1/\mu_2$, cf. Cox [1]. The Laplace transform of a Coxian-2 density is the ratio of a polynomial of degree at most one to a polynomial of degree two. The latter property defines the class of K_2-distributions, cf. Cohen [2]. In fact, every K_2-density is equivalent to a Coxian-2 density. Denoting by $W_q(.)$ the probability distribution of the waiting time of a customer in the stationary situation and assuming service in order of arrival it follows from general results in Cohen [2], p 324, that for Coxian-2 service

$$1 - W_q(t) = \sum_{i=0}^{2} a_i e^{-\eta_i t} \quad \text{for all } t \geq 0 \tag{2}$$

where $0 < \eta_1 < \min(\mu_1, \mu_2) < \eta_2$ are the roots of the equation

$$x^2 - (\mu_1 + \mu_2)x + \mu_1\mu_2 - (\mu_1\mu_2 - (1-b)\mu_1 x)\int_0^\infty e^{-xt} dA(t) = 0 \tag{3}$$

and the constants a_i are given by

$$a_1 = \frac{\eta_2(\eta_1 - \mu_2)(\mu_1 - \eta_1)}{\mu_1\mu_2(\eta_1 - \eta_2)}, \qquad a_2 = \frac{\eta_1(\eta_2 - \mu_2)(\mu_1 - \eta_2)}{\mu_1\mu_2(\eta_2 - \eta_1)} \tag{4}$$

Here A(t) denotes the probability distribution function of the interarrival time. Denoting by ρ the quotient of the mean service time and the mean interarrival time, the condition $\rho<1$ is required in order to guarantee that $\lim_{n\to\infty} P(W_n \leq x) = W_q(x)$ exists and is a proper probability distribution where W_n is the delay in queue of the n-th arrival. For the special case of exponential service we actually have $1 - W_q(x) = ae^{-\eta x}$ where η is the solution of the equation

$$x - \mu_1 + \mu_1 \int_0^\infty e^{-xt} dA(t) = 0 \quad \text{and} \quad a = \frac{\mu_1 - \eta}{\mu_1}$$

The usefulness of the tractable analytical solution (2) is not restricted to C_2-service. This solution can also be used to calculate simple and accurate approximations for the general GI/G/1 queue by using extrapolation with respect to the squared coefficient of variation of the service time, cf. Tijms [3]. The result (2) was obtained in Cohen [2] in a much more general framework by rather deep arguments using complex function analysis. In view of the practical importance of the result (2), it seems of some interest to give an elementary proof that can be used in introductory courses on queueing models. The elementary proof is based on the observation that a Coxian-2 distributed service time can be represented as a random sum of exponentially distributed service phases with the *same* means. This observation allows us to use a simple embedded Markov-chain analysis. In section 2 it is shown that the waiting-time distribution for the GI/G/1 model can be obtained by analyzing a special batch-arrival $GI^X/M/1$ model. The embedded Markov-chain analysis is given in section 3.

2. The analysis for the waiting-time distribution.

To give an elementary derivation of the waiting-time distribution we need the following well-known result. Suppose that $X_1, X_2, ..$ are independent random variables with a common exponential distribution with mean $1/\mu$ and that the random variable N has a geometric distribution $P\{N=k\} = p(1-p)^{k-1}$ for k = 1, 2, ... If the random variable N is independent of the sequence $\{X_n\}$ then the random sum $\sum_{i=1}^{N} X_i$ is exponentially distributed with mean $1/(p\mu)$.

Consider now the C_2-distributed service time S as in (1). Depending on whether $\mu_1 \geq \mu_2$ or $\mu_2 > \mu_1$ we represent S_1 or S_2 as a geometrically distributed sum of independent exponentials where the parameters p and μ of this representation are $p=\mu_2/\mu_1$, $\mu=\mu_1$ respectively $p=\mu_1/\mu_2$ and $\mu=\mu_2$. For the case $\mu_1 \geq \mu_2$ it can be said that a customer represents a batch of independent service phases where the batch-size distribution $\{g_k\}$ is given by

$$g_k = \begin{cases} 1-b & \text{for } k = 1 \\ bp(1-p)^{k-2} & \text{for } k = 2, 3, .. \end{cases} \tag{5}$$

with $p=\mu_2/\mu_1$. Here the service time of each phase is exponentially distributed having mean $1/\mu$ with $\mu=\mu_1$. For the case of $\mu_2>\mu_1$ we get the same representation (5) provided we replace b by $1-(1-b)p$ and take $p=\mu_1/\mu_2$ and $\mu=\mu_2$. In other words the $GI/C_2/1$ queue can be studied as a special $GI^X/M/1$ queue with batch-size distribution (5). Irrespective of the special form of g_k, we can define the following embedded Markov chain

X_n - the number of uncompleted exponential service phases present in the system just prior to the arrival epoch of the n-th customer.

Since $\rho<1$ the Markov chain $\{X_n : n=1,2,..\}$ is positive-recurrent. Moreover the chain is aperiodic. Let us denote its one-step transition probabilities by p_{ij} and its n-step transition probabilities by $p_{ij}^{(n)}$. Then from Markov-chain theory we know that the limiting distribution

$$\pi_j = \lim_{n\to\infty} P(X_n=j) = \lim_{n\to\infty} p_{ij}^{(n)}, \qquad j = 0, 1, ..$$

exists and is independent of the initial state i. Moreover, $\{\pi_j\}$ is the unique nonnegative solution of the system of linear equations

$$x_j = \sum_{i=0}^{\infty} x_i p_{ij} \quad , \; j = 0, 1, .. \tag{6}$$

together with the normalization equation $\sum_{j=0}^{\infty} x_j = 1$.
The limiting distribution $\{\pi_j\}$ leads directly to the waiting-time distribution of a customer arriving in the stationary situation. Since the phases are independent exponentials with a common mean $1/\mu$ and because of the memoryless property of the exponential distribution the conditional delay in queue of a customer finding upon arrival j phases in front of him has an Erlang-j distribution. Hence

$$W_q(t) = \sum_{j=1}^{\infty} \pi_j (1 - \sum_{k=0}^{j-1} e^{-\mu t}(\mu t)^k/k!) \qquad \text{for all } t\geq 0 \tag{7}$$

In the next section it will be shown that for certain constants α, β, τ_1 and τ_2

$$\pi_j = \alpha\tau_1^{\,j} + \beta\tau_2^{\,j} \qquad \text{for } j = 0, 1, \ldots$$

Then, together with (7) we get

$$1 - W_q(t) = \frac{\alpha\tau_1}{1-\tau_1} e^{-(1-\tau_1)\mu t} + \frac{\beta\tau_2}{1-\tau_2} e^{-(1-\tau_2)\mu t} \tag{8}$$

At the end of section 3 we will see that (8) and (2) coincide.

3. Embedded Markov-chain analysis.

In this section we will first prove that there exist numbers τ_1, τ_2, α and β such that $x_j = \alpha\tau_1^{\,j}+\beta\tau_2^{\,j}$ for j=0,1,.. satisfies simultaneously the system of linear equations (6) and the normalization equation. Subsequently we will show

that this solution coincides with the steady-state distribution $\{\pi_j\}$.
Put for abbreviation

$$k_n = \int_0^\infty e^{-\mu t} \frac{(\mu t)^n}{n!} dA(t), \qquad n = 0, 1, \ldots$$

Then, we have obviously

$$p_{ij} = \sum_{\ell=(j-i)^+}^{\infty} g_\ell k_{i+\ell-j} \qquad \text{for } j = 1, 2, ..; \ i = 0, 1, .. \tag{9}$$

Here $(j-i)^+ = \max(0, j-i)$ and $g_0=0$. Note that the formulae (9) only apply for $j\neq 0$ and that $p_{i0}= 1 - \sum_{j\neq 0} p_{ij}$. With (9) we can write the equations in (6) as

$$x_j = \sum_{i=0}^{\infty} x_j \sum_{\ell=(j-i)^+}^{\infty} g_\ell k_{i+\ell-j}, \qquad \text{for } j = 1, 2, .. \tag{10}$$

Now define for any solution $\{x_j\}$ of (10)

$$q_j = \sum_{i=0}^{j} x_i g_{j-i}, \qquad \text{for } j = 0, 1, .. \tag{11}$$

If we take $x_j=\pi_j$ for all $j\geq 0$, q_j represents the probability that j uncompleted phases are present in the system just after the arrival of a customer. Using that $g_k = (1-p)g_{k-1}$ for $k\geq 3$ (see (5)) we find

$$q_j= g_1x_{j-1}+ g_2x_{j-2}+ (1-p)\sum_{i=0}^{j-2} x_i g_{j-1-i} - (1-p)x_{j-2}g_1 \quad , \quad j = 2, 3, ..$$

and so we obtain the recursion relation

$$q_j = (1-p)q_{j-1} + (1-b)x_{j-1} +(b+p-1)x_{j-2}, \quad \text{for } j = 2, 3, \ldots \tag{12}$$

Now we rewrite (11) in a more convenient form by changing the order of summation. By summing over the diagonals $n = i+\ell$ we get

$$x_j = \sum_{n=j}^{\infty} \sum_{i=0}^{n} x_i g_{n-i} k_{n-j}, \qquad \text{for } j = 1, 2, \ldots \tag{13}$$

Next by inserting (11) in (13) and using the recursion (12) we find

$$x_j = (1-p)x_{j-1}+ (1-b)\sum_{n=j}^{\infty} x_{n-1}k_{n-j}+ (b+p-1)\sum_{n=j}^{\infty} x_{n-2}k_{n-j} \quad , \ j = 2, 3, .. \tag{14}$$

In fact the equations (14) are equivalent to the equations (13) in which the equation corresponding to j=1 is omitted. Denote by $\hat{A}(s) = \int_0^\infty e^{-st} dA(t)$ the Laplace-Stieltjes transform of the interarrival-time distribution function

A(t). Then, by substitution, we find that the system of equations (14) allows solutions of the form $x_j = \tau^j$ (j = 0, 1, ..) if τ satisfies the equation

$$\frac{z^2 - (1-p)z}{(1-b)z + b+p-1} = \hat{A}(\mu(1-z)). \tag{15}$$

In the sequel we assume that both $b+p-1\neq0$ and $b\neq0$. If either condition is not satisfied the $GI/C_2/1$ queue reduces to the well-studied GI/M/1 queue in which case (15) has a unique solution $\tau>0$ in the interval (-1, 1). Using the condition $\rho<1$ it is a matter of routine analysis to show that (15) has two real roots $\tau_1<\tau_2$ in the open interval (-1, 1) if $b+p-1\neq0$. In case $b+p-1>0$ we have $\tau_1<0$,while $\tau_1>0$ in case $b+p-1<0$. In both cases $\tau_2>1-p$. Once we know that (14) allows solutions $x_j=\tau_1{}^j$ and $x_j=\tau_2{}^j$ it is immediate that any linear combination $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$ (j=0,1,..) is also a solution of the system (14).

We now conjecture $\pi_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$, j=0,1,.., with α and β still to be specified. To prove this conjecture we must determine α and β such that $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$, j=0,1,.., is a solution of system (6). Because the equations (6) with j=2,3,.. are equivalent with (14) and $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$ is a solution of (14), we have that the proposed solution satisfies the equations (6) for j=2,3,.. irrespective of the values of α and β. So we still have to prove that there exist an α and a β such that $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$ is also a solution of both the remaining equation of system (6)

$$x_1 = \sum_{i=0}^{\infty} x_i p_{i1} \tag{16}$$

and the normalization equation $\sum_{j=0}^{\infty} x_j = 1$.
We do this by substituting the solutional form $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$ into these equations. With respect to (16) this leads after some algebra (using (13) with j=1) to the following equation

$$\alpha\tau_1 + \beta\tau_2 = bp\left\{ \frac{\alpha\hat{A}(\mu(1-\tau_1))}{\tau_1+p-1} + \frac{\beta\hat{A}(\mu(1-\tau_2))}{\tau_2+p-1} - \hat{A}(p\mu)\left(\frac{\alpha}{\tau_1+p-1} + \frac{\beta}{\tau_2+p-1} \right)\right\} \tag{17}$$

Now we use the fact that τ_1 and τ_2 satisfy (15), so we can replace $\hat{A}(\mu(1-\tau_i))$, (i=1,2) by the left side of (15). After some simplifications this gives

$$\alpha\tau_1 + \beta\tau_2 = \alpha\tau_1 + \beta\tau_2 - bp\hat{A}(p\mu)\left(\frac{\alpha}{\tau_1+p-1} + \frac{\beta}{\tau_2+p-1} \right)$$

So we find, $\alpha/(\tau_1+p-1) + \beta/(\tau_2+p-1) = 0$. Further, substitution of the proposed solution in the normalizing equation $\sum x_j = 1$ gives, $\alpha/(1-\tau_1) + \beta/(1-\tau_2) = 1$. From these two linear equations in α and β we find

$$\alpha = \frac{(1-\tau_1)(1-\tau_2)(1-p-\tau_1)}{p(\tau_2-\tau_1)} \qquad \beta = \frac{(1-\tau_1)(1-\tau_2)(1-p-\tau_2)}{p(\tau_1-\tau_2)}. \tag{18}$$

Taking these values for α and β the solution $x_j=\alpha\tau_1{}^j+\beta\tau_2{}^j$, $j=0,1,\ldots$, satisfies all the balance equations in (6) together with the normalization equation $\sum x_j=1$. So it remains to show that $x_j=\pi_j$ for all $j\geq 0$. For completeness we include the proof of this. The series $\sum x_j$ converges absolutely because both $|\tau_1|$ and $|\tau_2|$ are less than one. Thus, using that for the ergodic Markov chain $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$, we find by iterating the balance equations (6) that $x_j = \sum_{i=0}^{\infty} x_i p_{ij}^{(n)}$, for all $n\geq 1$. Next, letting $n\to\infty$, we get $x_j = \sum_i x_i \pi_j$, using the bounded convergence theorem. This gives $x_j=c\pi_j$ for all $j\geq 0$ with $c=\sum x_j$. But from the earlier analysis we know that $\sum x_j=1$, and so $x_j=\pi_j$ for all j. Finally, we show that formulae (2) and (8) coincide. First we remark that the transformation $\mu(1-z)=x$ in (15) leads to (3). This gives $\mu(1-\tau_1)=\eta_2$ and $\mu(1-\tau_2)=\eta_1$. So, also using (18), we can express (8) in η_1 and η_2 and after some algebra this expression turns out to be equal to (2).

Acknowledgement.

We would like to thank prof. dr. H. C. Tijms for his valuable guidance during the course of this work. He suggested many improvements which made the text much more readable.

References.

[1] Cox, D. R. (1955). A use of complex probabilities in the theory of stochastic processes, Proc. Camb. Phil. Soc., 51, 313-319.

[2] Cohen, J. W. (1982). The Single Server Queue, 2nd ed., North-Holland, Amsterdam.

[3] Tijms, H. C. (1986). Stochastic Modelling and Analysis, Wiley, New York.