1986

The Presence of Exponentiality in Entropy Maximised M/GI/1 Queues

Vernon J. Rego
Purdue University, rego@cs.purdue.edu

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
86-620

THE PRESENCE OF EXPONENTIALITY IN
ENTROPY MAXIMISED M/GI/1 QUEUES

Vernon Rego
Wojciech Szpankowski

CSD-TR-620
August 1986

Abstract

It is known that entropy maximised M/GI/1 queues yield queue length distributions that are geometric when the constraints involve only the first two moments of the service time distribution. By proving that geometric queue lenth is equivalent to exponential service in M/GI/1 queues, we show that using the entropy maximisation procedure with only two service time moments is equivalent to using exponentially distributed service times. Thus, by identifying the parameter of the equivalent service time distribution, we can compute such entropy maximised solutions via classical M/M/1 theory. In case an additional constraint is used to preserve the probability of an empty queue, the service time distribution becomes a mixure of two distributions, one of which is exponential, and the other an impulse function of unit mass at the origin. In this case, except for the probability of an empty queue, the remaining queue length probabilities follow a geometric sequence. In either case, our results demonstrate the presence of exponentiality in the service times of entropy maximised M/GI/1 queues.

1. Introduction

Within the past decade, there has developed a considerable amount of interest in the application of information-theoretic methods in queueing theory. In particular, we refer to the increasing interest in the use of maximum entropy methods in obtaining equilibrium distributions for stochastic processes, such as queue lengths in single-server, multi-server, and networks of queues, and performance distributions in computer systems performance evaluation. The origins of the general method lie in the extension of Laplace's principle of insufficient reason, by E.T. Jaynes [1]. The result of the extension was a non-linear programme with linear constraints, now often used to obtain an equilibrium probability distribution given only its expectation.

In one of the earliest applications of the *entropy maximisation* (EM) method, Benes [2] demonstrated an equivalence between a Markov process and the maximum entropy principle. Later, Ferdinand [3, 4] refined and applied EM to analyze systems, and in particular, the machine interference problem. Since then, EM has been applied in various situations, such as in queueing networks by Kouvatsos [5], in the analysis of computer systems performance by Bard [6], and in the analysis of single and infinite server queues by Shore [7], and El-Affendi and Kouvatsos [8].

The organisation of this paper is as follows. In section 2 is presented a very brief review of the entropy maximisation method (for more details the reader is referred to [1], [7] and [8]), along with key results from Shore [7], and El-Affendi and Kouvatsos [8]. In section 3 we also remark on the use of EM, and the nature of this paper's contribution. In section 3 is presented a set of main results which are used to characterise exponential service in M/GI/1 queues. In essence, we show that there is a unique relationship between service times that are exponential and queue lengths that are geometric. In section 4, we present a main result, namely, that in applying entropy maximisation to an M/GI/1 queue with only the first two moments of the service time distribution as constraints, we are in actuality assuming exponentially distributed service times. Thus, we show that Shore's results on the M/GI/1 queue length distribution is equivalently obtained via classical M/M/1 queueing theory. Additionally, we show that in the presence of another constraint involving the probability of an empty queue, the service time distribution still persists in having a strongly exponential component.

2. The Entropy Maximization Method

Consider a system X that has a finite or countable set of states S. Let p_j denote the probability that the system X is in state j, $j \in S$. In addition, there is information available about the system X in the form of $(m+1)$ constraints,

$$\sum_{j \in S} p_j = 1 \tag{1}$$

$$\sum_{j \in S} f_k(j)\, p_j = v_k, \qquad 1 \le k \le m < \infty \tag{2}$$

where $\{v_k\}$ is a set of expectations defined through appropriate functions $\{f_k(j);\ j \in S\}$ of system states. Since $(m+1) < |S|$ in general, there will usually always be an infinite number of distributions $\{p_j\}$ that satisfy the constraints (1) and (2). The principle of entropy maximisation (EM) states that the least biased distribution satisfying (1) and (2) is the one that maximises Shannon's entropy,

$$H(\mathbf{p}) = -\sum_{j \in S} p_j \ln p_j \tag{3}$$

where $\mathbf{p} = (p_1, p_2, \cdots)$ is the vector of equilibrium state probabilities for the system X.

It is generally the case (see for example, [7-9]) that $m = 1$, and consequently only two constraints are used. In this case, the constraints are simply $\sum_{j \in S} p_j = 1$, and $\sum_{j \in S} f(j) p_j = v$ for some convenient function f defined over S. The distribution obtained by EM is given by (see [9])

$$p_j = \frac{1}{\phi(\beta_0)} e^{-\beta_0 f(j)}, \quad j \in S \tag{4}$$

where $\phi(\beta) = \sum_{j \in S} e^{-\beta f(j)}$, with β_0 as the unique solution of the equation

$$\frac{d \ln \phi(\beta)}{d \beta} = - \nu \tag{5}$$

In general, it is usually impossible to solve (5) or its m-dimensional counterpart (using m constraints and the normalisation) explicitly, so that a closed-form solution can be obtained in terms of ν or the set of expectations $\{ \nu_1, \cdots, \nu_m \}$. Thus, the solution is usually obtained in terms of Lagrangian multipliers ([7], [9]). In specific cases, solutions can be obtained that reduce to surprisingly simple forms. In this paper we address one such set of results from [7] and [8] that yields geometric or geometric-type distributions for the equilibrium queue length of the M/GI/1 queue. In the following paragraphs, we present results from Shore [7] and El-Affendi and Kouvatsos [8] so that we may directly refer to them later.

The Shore Theorem:

For an M/GI/1 queueing system with equilibrium queue length distribution $\{q_n\}$, the result of maximising the entropy of $\{q_n\}$ subject to the single known constraint $\sum_{n=1}^{\infty} n\, q_n = L$ and the normalisation constraint $\sum_{n=1}^{\infty} q_n = 1$, is

$$q_n = \frac{1}{L+1} \left(\frac{L}{L+1} \right)^n , \quad n \geq 0 . \tag{6}$$

∎

The El-Affendi and Kouvatsos Theorem:

In an M/GI/1 queueing system with customer arrival rate λ, mean service time $\frac{1}{\mu}$, squared coefficient of variation of service time distribution given by C_s^2, and expected value of the equilibrium queue length distribution $\{r_n\}$ given by the Pollaczek-Khinchin formula

$$L = \rho + \frac{\rho^2 (1 + C_s^2)}{2(1-\rho)} , \tag{7}$$

the maximum entropy solution subject to the constraints $\sum_{n=1}^{\infty} n\, r_n = L$, $\sum_{n=1}^{\infty} r_n = 1$, and $p_0 = 1 - \rho$, is given by

$$r_n = (1-\rho)\,\frac{1}{(y_s+1)}\left\{\frac{\rho(y_s+1)}{1+\rho y_s}\right\}^n, \tag{8}$$

for $n \geq 1$, with $\rho = \dfrac{\lambda}{\mu} < 1$, and $y_s = \dfrac{(C_s^2-1)}{2}$. ■

Remarks on the use of EM and contributions of this paper

The aim of this paper is to use simple probabilistic arguments to show that there is a strong equivalence between using EM and assuming exponentiality. We focus on the Shore Theorem (from [7]), an important application of EM, to show that the result so obtained (i.e., equation (6)) is just as easily obtained via classical queueing theory. Next, we carry our arguments over to the the El-Affendi and Kouvatsos theorem (from [8]), to show that equation (8) really describes a geometric-type distribution. In an attempt to *preserve* the value of the probability of an empty queue, r_0, these authors use the notion of *prior information* in executing the maximum entropy procedure. In effect what they achieve is a preservation of $r_0 = 1-\rho$ at the expense of scaling the rest of the distribution. Nevertheless, the rest of the distribution *still follows a geometric sequence*. The most important consequence of our findings is that the EM method *will fail* to give good results if the system being modelled exhibits non exponential behaviour, since it is well known that equilibrium distributions can behave very differently from geometric ones (for example, see Neuts [14], or in another treatment [16]).

The EM technique is an elegant and powerful technique, but only when properly used. We wish to draw special attention to the fact that the method is most often applied subject to only the normalisation and the mean value constraints. Our results bear very strongly on this aspect of its application. If from available data we took pains to estimate the second and higher order moments of random variables and use all these as constraints, there is little doubt that the EM method may well approximate the actual distribution. However, second and higher order moments become increasingly difficult to estimate, and as a consequence, are conspicuously absent from the set of constraints of most EM models. This, we feel, is one major shortcoming of the EM application. Additionally, even if one did obtain higher order estimates of moments, their use as constraints adds considerable complexity to the solution. Thus, while a conscientious application of EM is undoubtedly good in principle, in practice, there is much to be worked out, for sound solutions to be obtained.

Shore [7] points out that equilibrium queue length distributions obtained via EM possess the property that they are monotonically decreasing, for a wide class of M/GI/1 systems. He further argues that since this is a feature typically seen in few-moment approximations and *not* many-moment approximations, it will suffice to use only a few moments in order to obtain a low mean-squared-error between the actual and the entropy maximised queue length distributions. Our results imply that using only the first two moments of the service time distribution (or equivalently, only the first moment of the queue length distribution) along with the normalisation constraint yields a queue-length distribution that can only come from assuming that the system is an M/M/1 system. Thus if the service time distribution is far from geometric, we can conclude that the queue length distribution will also be far from geometric. In this sense, the geometric queue length distributions that EM yields can be exceedingly poor, and consequently such use of EM methods can often lead to deplorable performance distributions.

3. On Characterising Exponential Service

In this section we present some key results, summarised as follows. In the first part we present a brief sequence of results that characterise the M/M/1 queue. These are used to prove a main theorem, which effectively states that for a single server queue, the equilibrium queue length distribution $\{p_n\}_{n=0}^{\infty}$ is a geometric distribution if and only if the queue is a *pure* M/M/1 queue (i.e., *pure* in the sense that the customer interarrival time and service time distributions are strictly exponential distributions). If we modify this slightly and say that given $p_0 = 1-\rho$ (as *prior information*), the sequence $\{p_n\}_{n=1}^{\infty}$ is a geometric sequence, then we obtain the following interesting result. If $\{p_n\}$ is a *tainted* geometric distribution (i.e., there is a contaminating influence at the origin, but the rest of the distribution follows a geometric sequence), then the service time distribution becomes a mixture distribution, with one component that is exponential, and another component that is a unit impulse mass at the origin. We refer to the latter distribution as a *tainted* exponential distribution (i.e., there is a contaminating influence at the origin) and later give a motivation for reducing it to a pure exponential distribution provided a certain requirement is met. We call such an M/GI/1 system with tainted exponential service (or equivalently, with tainted geometric queue lengths) a *tainted* M/M/1 queue.

The first lemma, which we only state, is due to Engel and Zijlstra. The proof may be found in [10]. Consider a homogeneous Poisson process with parameter λ, $\lambda > 0$, and let $N(Y)$ be the number of points in the time-interval $[0, Y)$, with Y a non-negative random variable. Note that the range of $N(Y)$ is $\{0, 1, 2, \cdots\}$.

LEMMA 1

Let Y and Z be independent non-negative random variables with distribution functions F and G respectively. Then F is equal to G if $N(Y)$ and $N(Z)$ have the same probability distribution. ■

The above lemma is considered a variant of the moment problem (see Feller [11], chapter 7). In effect, Lemma 1 says that the Poisson process maps each probability distribution on $[0,\infty)$ onto a discrete probability distribution on $\{0, 1, \cdots\}$, and this mapping is one-to-one. In the next result, we show that if Y is exponentially distributed, then the number of Poisson generated points in $[0, Y)$ must be a geometric random variable.

LEMMA 2

The random variable $N(Y)$ has a geometric distribution with parameter $\dfrac{\lambda}{\mu + \lambda}$ if and only if Y is exponentially distributed with parameter μ, $\mu > 0$.

Proof.

Let Y have an exponential distribution with density

$$h(y) = \mu e^{-\mu t}, \quad t \geq 0 \tag{9}$$

The random variable $N(Y)$ takes the value j with probability

$$Pr(N(Y) = j) = \int_0^\infty e^{-(\lambda+\mu)t} \frac{\mu(\lambda t)^j}{j!} dt = \left[\frac{\mu}{\mu+\lambda}\right] \left[\frac{\lambda}{\mu+\lambda}\right]^j, \quad j \geq 0$$

which is a geometric distribution with parameter $\left[\frac{\lambda}{\mu+\lambda}\right]$.

Conversely, if $N(Y)$ is geometrically distributed with parameter $\frac{\lambda}{\mu+\lambda}$, a direct application of Lemma 1 yields Y as an exponential random variable with parameter μ. ∎

Consider an M/GI/1 queueing system with arrival rate λ, $\lambda > 0$, and a service-time distribution $B(\cdot)$. The probability that j customers arrive during an arbitrary customer's service is given by

$$k_j = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^j}{j!} dB(t) \tag{10}$$

thus yielding a distribution $\{k_j\}_{j=0}^\infty$ on the nonnegative integers. Assuming a stable queue, let $\{p_j\}_{j=0}^\infty$ denote the equilibrium queue length distribution of the M/GI/1 queue. We now show that if $\{k_j\}_{j=1}^\infty$ is a geometric sequence then $\{p_j\}_{j=1}^\infty$ must be a geometric sequence and vice-versa. Note that by Lemmas 1 and 2, $\{k_j\}_{j=0}^\infty$ is a geometric sequence if and only if the service time distribution of customers is exponential. The fact that $k_j = Pr(N(Y)=j)$ for $j \geq 0$, for $\{k_j\}_{j=0}^\infty$ a geometric distribution (or equivalently, Y an exponential random variable) is critical in what follows.

THEOREM 1

For a stable queueing system, the sequence $\{k_j\}_{j=1}^\infty$ is geometric if and only if the sequence

$\{p_j\}_{j=I}^{\infty}$ is geometric, where $I \in \{0, 1\}$.

Proof.

We first prove the theorem for the case $I = 1$. For an M/GI/1 queue to be stable, it is sufficient that [15]

$$\rho = \sum_{j=1}^{\infty} j\, k_j < 1 \tag{11}$$

If $\{k_j\}_{j=1}^{\infty}$ is a geometric sequence, with $k_j = \alpha\beta^{j-1}$, $0 < \alpha, \beta < 1$, $(\alpha+\beta) < 1$, $j \geq 1$, then ρ can be expressed as

$$\rho = \sum_{j=1}^{\infty} j\, \alpha\beta^{j-1} = \frac{\alpha}{(1-\beta)^2} \tag{12}$$

which implies that $\alpha < (1-\beta)^2$ is required for the system to be stable. Since $\{k_j\}_{j=0}^{\infty}$ is a distribution, we require that

$$k_0 = 1 - \frac{\alpha}{1-\beta} \tag{13}$$

Using Lemma 2 from [12], and a standard M/GI/1 recurrence equation (see [12], or [16])

$$k_0 p_{n+1} = p_n - p_0 k_n - \sum_{m=1}^{n} p_m k_{n-m+1}, \qquad n \geq 0, \tag{14}$$

we obtain the linear recurrence

$$k_0 p_{n+1} + (\alpha - \beta k_0 - 1) p_n + \beta = 0 \tag{15}$$

which is valid for $n \geq 1$. Equation (15) has the characteristic equation [13]

$$k_0 z^2 + (\alpha - \beta k_0 - 1) z + \beta = 0 \tag{16}$$

which possesses the roots $z = 1$, and $z = \dfrac{\beta}{k_0}$. Using appropriate boundary conditions, we obtain

$$p_n = \rho \left[\frac{k_0}{\beta} - 1 \right] \left[\frac{\beta}{k_0} \right]^n, \qquad n \geq 1 \tag{17}$$

with $p_0 = 1 - \rho$. It is clear from (17) that with k_0 fixed, and $\{k_j\}_{j=1}^{\infty}$, we obtain $\{p_j\}_{j=1}^{\infty}$ as a geometric sequence.

Conversely, assume that (17) holds, and $p_0 = 1 - \rho$ is given. For convenience, we write $p_n = \gamma \delta^{n-1}$ for $n \geq 1$. From (17) we see that $p_1 = \rho\,(1 - \frac{\beta}{k_0})$, which can be simplified to give

$$p_1 = (1-\rho)\,(\frac{1}{k_0} - 1) \tag{18}$$

from which k_0 may be obtained. Using the recurrence equation (14), we obtain

$$k_n\,(p_0 + p_1) = p_n - k_0 p_{n+1} - \sum_{m=2}^{n} \gamma \delta^{m-1} k_{n-m+1} \tag{19}$$

and after some algebra,

$$k_n = \frac{\delta p_0}{(p_0 + p_1)} k_{n-1}, \qquad n > 1 \tag{20}$$

with $k_0 = \dfrac{1-\rho}{1-\rho+p_1}$. On making the substitutions $\gamma = \rho\,(k_0/\beta - 1)$ and $\delta = \beta/k_0$, we obtain $k_n = \alpha \beta^{n-1}$, for $n \geq 1$. Thus, we have that $\{k_j\}_{j=1}^{\infty}$ is a geometric sequence.

Next, we must prove the theorem for the case $l = 0$. If we assume that $\{k_j\}_{j=0}^{\infty}$ is geometric, then $k_j = a\,b^j$ for $j \geq 0$. On comparing with the sequence $\{k_j\}_{j=1}^{\infty}$, we see that this means $k_0 = a$, $\alpha = a\,b$, and $\beta = b$. Using this in (17), we obtain

$$p_n = \rho \left[\frac{a}{b} - 1 \right] \left[\frac{b}{a} \right]^n, \qquad n \geq 1, \tag{21}$$

with $p_0 = 1 - \rho$, and $a + b = 1$. But in this case, $\rho = \dfrac{1-a}{a}$, thus forcing $p_n = (1-\rho)\rho^n$, for $n \geq 0$, which of course means that $\{p_j\}_{j=0}^{\infty}$ is a geometric distribution.

Conversely, assume that $p_n = (1-\rho)\rho^n$, for $n \geq 0$. In this case, we obtain $p_0 = 1 - \rho$, $\gamma = (1-\rho)\rho$, and $\delta = \rho$. Using this in (20) we obtain $k_n = \dfrac{\rho}{1+\rho}$ for $n > 1$. The value of k_0 is obtained from $k_0 = \dfrac{1-\rho}{1-\rho+p_1} = \dfrac{1}{1+\rho}$, which necessarily means that $k_1 = \left[\dfrac{1}{1+\rho}\right]\left[\dfrac{\rho}{1+\rho}\right]$. It follows that $\{k_j\}_{j=0}^{\infty}$ is a geometric distribution. ■

In the next theorem, which ties the results of this section together, we show that the M/M/1 queue arises as a very special case of the M/GI/1 queue. In particular, the service time random variable in an M/GI/1 queue has a pure exponential distribution if and only if the equilibrium queue length random variable has a pure geometric distribution. Correspondingly, the service time random variable in an M/GI/1 queue has a tainted exponential distribution if and only if the equilibrium distribution of the queue length random variable has a tainted geometric distribution.

THEOREM 2

Consider a stable M/GI/1 queueing system with customer arrival rate λ, and service time distribution $B(\cdot)$, with finite first and second moments $\frac{1}{\mu}$ and σ^2, respectively. Let $\{p_j\}_{j=0}^{\infty}$ be the equilibrium queue length distribution of the queue, $\mu > 0$, and $\rho = \frac{\lambda}{\mu} < 1$.

Pure M/M/1 Queueing System

The following statements are equivalent:

(I) $B(\cdot)$ is a pure exponential distribution with parameter μ.

(II) The distribution $\{k_j\}_{j=0}^{\infty}$ is geometric, i.e., $k_j = a\, b^j$, for $j \geq 0$.

(III) The distribution $\{p_j\}_{j=0}^{\infty}$ is geometric, i.e., $p_j = \left(\frac{2a-1}{a}\right)\left(\frac{1-a}{a}\right)^j$, for $j \geq 0$.

Tainted M/M/1 Queueing System

Let α and β be given, such that $(\alpha+\beta) < 1$, and $\alpha > 0$, $\beta > 0$. Define $k_0 = 1 - \frac{\alpha}{1-\beta}$.

The following statements are equivalent:

(IV) $B(\cdot)$ is a tainted exponential distribution, explicitly given as

$$B(t) = \left(\frac{\beta+k_0-1}{\beta}\right)\delta_0(t) + \left(\frac{1-k_0}{\beta}\right)^2 \mu\, e^{-(1-k_0)\mu t/\beta} \tag{22}$$

where $\delta_0(\cdot)$ is the impulse function with unit mass, defined as $\delta_0(t) = \infty$ when $t = 0$,

$\delta_0(t)=0$ when $t>0$, and $\int_0^\infty \delta_0(t)\,dt=1$.

(V) The distribution $\{k_j\}_{j=0}^\infty$ is a tainted geometric distribution, with $k_j=\alpha\beta^j$, for $j \geq 1$.

(VI) The distribution $\{p_j\}_{j=0}^\infty$ is a tainted geometric distribution, with $p_j=\rho\left[\frac{k_0}{\beta}-1\right]\left[\frac{\beta}{k_0}\right]^j$,

for $j \geq 1$.

Proof.

Using Lemma 1, we see that (II) follows directly from (I). Using Theorem 1, it is clear that (III) follows from (II). From Theorem 1 it is also clear that (I) is a consequence of (III). Thus, the pure M/M/1 part of the theorem is proved.

From Theorem 1, we see that (V) implies (VI) and vice-versa. We now show that (IV) follows from (VI). Assume that (VI) holds. Then the generating function of the queue length distribution is given by

$$\begin{aligned} L(z) = \sum_{n=0}^{\infty} p_n z^n &= (1-\rho) + \sum_{n=1}^{\infty} \rho \left[\frac{k_0-\beta}{\beta}\right]\left[\frac{\beta z}{k_0}\right]^n \\ &= (1-\rho) + (1-\rho) \sum_{n=1}^{\infty} \left[\frac{1-k_0}{\beta}\right]\left[\frac{\beta z}{k_0}\right]^n \\ &= (1-\rho)\left[\frac{1-\dfrac{\beta+k_0-1}{k_0}z}{1-\dfrac{\beta}{k_0}z}\right]. \end{aligned} \tag{23}$$

On the other hand, the Pollaczek-Khinchin transform equation independently yields

$$L(z) = \frac{B^*(\lambda(1-z))(1-\rho)(1-z)}{B^*(\lambda(1-z))-z}, \tag{24}$$

where $B^*(s)$ is the Laplace-Stieltjes transform of the service time distribution $B(\cdot)$. On comparing (23) with (24), and solving for $B^*(\lambda(1-z))$, using $s=\lambda(1-z)$, we obtain

$$B^*(s) = \frac{\beta+k_0-1}{\beta} + \frac{\left[\frac{1-k_0}{\beta}\right]^2 \mu}{\left[\frac{1-k_0}{\beta}\right]\mu + s} \tag{25}$$

which, upon inversion, yields the service time density in (IV). Thus, (IV) follows from (VI).

Finally, we are left to prove that (IV) implies (V). We compute k_j, $j \geq 0$, using the definition in (10) and the distribution in (22), in a straightforward fashion. We obtain

$$k_0 = 1 - \kappa + \kappa \left[\frac{\kappa\mu}{\lambda + \kappa\mu} \right]$$

where $\kappa = \frac{\alpha}{\beta(1-\beta)}$. Observe that the unit impulse function $\delta_0(\cdot)$ plays no part in the computation of k_j, $j \geq 1$. In computing the latter, we obtain

$$k_j = \frac{\kappa\rho^*}{\left[1+\rho^*\right]^2} \left[\frac{\rho^*}{1+\rho^*} \right]^{j-1}, \qquad j \geq 1$$

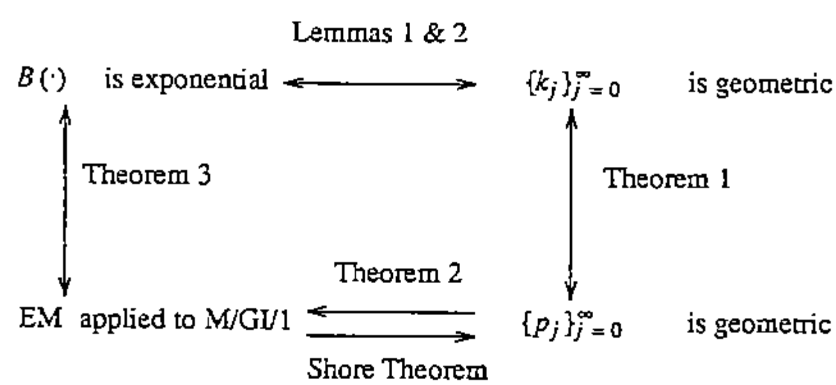
where $\rho^* = \frac{\lambda}{\mu^*}$, and $\mu^* = \kappa\mu$. Thus, we obtain $\{k_j\}_{j=1}^{\infty}$ as a geometric sequence, and (V) is shown to follow from (IV). ■

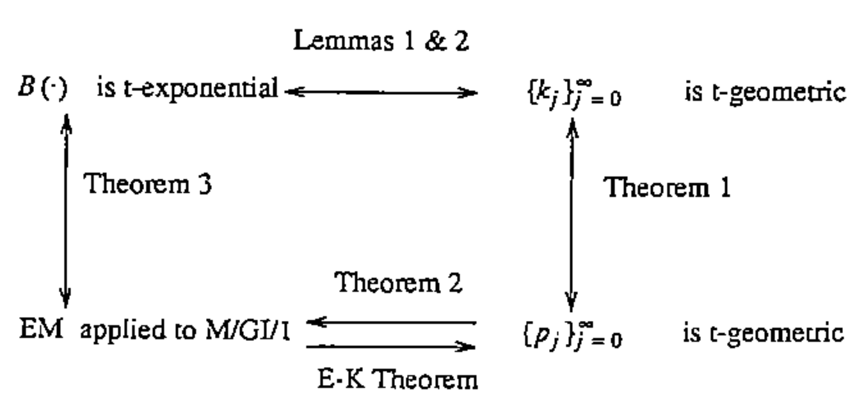
4. The Presence of Exponentiality in Entropy Maximisation

In this section we link our results to the theorems of Shore, and El-Affendi and Kouvatsos. Using the simple probabilistic derivations presented in the last section, we obtain Shore's maximum entropy queue length distribution in a very direct fashion (without resorting to a linear programme). We show that Shore's maximum entropy queue length distribution can be obtained if and only if the system being analysed is an M/M/1 system. A clear consequence of this fact is that the EM method can be expected to yield good approximations to performance distributions *if and only if* the system under observation is is *close* to a pure M/M/1 system, or a tainted M/M/1 system. In the latter case, the use of *prior information* to preserve the value of $p_0 = 1 - \rho$ in the El-Affendi and Kouvatsos theorem forces an otherwise pure M/M/1 system to become mildly tainted.

In effect, this section presents an alternate way to obtain an entropy maximised performance distribution for single server queueing systems, and it is apparent that these results possess several extensions. The technique can be extended to multi-server queues and networks of queues. In addition, the strong equivalence between exponential service and geometric queue length distributions implies that both, our results, as well as the EM method will work well in heavy traffic approximations, diffusion approximations, etc. If the EM method has worked well in approximating such processes (with geometric or tainted geometric equilibrium distributions) our results give an indication as to why this happens. These ideas are to be presented in greater detail in a forthcoming paper.

The following diagrams summarise the results presented thus far, and graphically explain the function of each result in proving the equivalence of EM and assuming exponentiality. We also present a final theorem that demonstrates the equivalence, and determines the parameters of the equivalent M/M/1 system.





The notation t-exponential is short for tainted exponential distribution, and the E-K theorem is the theorem of El-Affendi and Kouvatsos.

In the following result, we demonstrate an equivalence between maximising the entropy and assuming exponentiality of service time, of an M/GI/1 queue. The essence of the idea is as follows. Suppose that we assume that the queue length distribution of an M/GI/1 queue is geometric, with parameter ρ'. That is, $q_n = (1-\rho')(\rho')^n$, for $n \geq 0$. This implies that the average queue length is $L = \sum_{n=0}^{\infty} n\, q_n = \frac{\rho'}{1-\rho'}$, from where is obtained $\rho' = \frac{L}{L+1}$. Similarly, for the El-Affendi and Kouvatsos result, suppose that the steady-state distribution $\{r_n\}$ satisfies $r_0 = 1-\rho$, and $r_n = (1-\rho)\,\kappa \left\{\rho^*\right\}^n$, for $n \geq 1$. The constant κ can be computed directly from the normalisation $\sum_{n=0}^{\infty} r_n = 1$. In order to determine ρ^*, note that the average queue length is given by $\frac{(1-\rho)\,\kappa\rho^*}{(1-\rho^*)^2}$. On comparing this with the Pollaczek Khinchin formula (7), ρ^* can be obtained.

THEOREM 3 (*Equivalence Principle*)

Consider an M/GI/1 queueing system having arrival rate λ, service time mean $\frac{1}{\mu}$, service time distribution $B(\cdot)$, with $\lambda > 0$, $\mu > 0$, and $\rho = \frac{\lambda}{\mu} < 1$. Let C_s^2 denote the squared coefficient of variation of $B(\cdot)$, define $y_s = \frac{(C_s^2 - 1)}{2}$, and let L denote the equilibrium mean queue length.

(I) In the absence of *prior information* regarding q_0, the equilibrium distribution $\{q_n\}$ given in (6) can be obtained if and only if the customer service time distribution is pure exponential, with mean $\mu' = \frac{\lambda(L+1)}{L}$,

and

(II) in the presence of *prior information* that fixes $r_0 = 1-\rho$, the distribution $\{r_n\}$ given in (7) can be obtained if and only if the customer service time distribution is tainted exponential, of the form

$$f(t) = \left[\frac{y_s}{y_s+1}\right] \delta_0(t) + \left[\frac{1}{y_s+1}\right]^2 \mu e^{-\mu t/(y_s+1)} \tag{26}$$

Proof.

We begin by proving (I). From section 3, we know that a geometric queue length distribution results if and only if the service time distribution is exponential. Assume that the service time distribution is exponential, with mean $\mu' = \frac{\lambda(L+1)}{L}$. Since the interarrival time distribution is also exponential, we have an M/M/1 system, with traffic intensity

$$\rho' = \frac{\lambda}{\mu'} = \frac{L}{L+1}.$$

Clearly, the queue length distribution is necessarily geometric, of the form

$$\begin{aligned} p_n &= (1-\rho')\left[\rho'\right]^n, \quad n \geq 0 \\ &= \frac{1}{L+1}\left[\frac{L}{L+1}\right]^n, \quad n \geq 0 \end{aligned}$$

which is precisely the distribution (6) given by the Shore theorem. Conversely, if (6) is the queue length distribution of an M/GI/1 queue, with arrival rate λ, then Theorem 2 tells us that the service time distribution must be exponential, with parameter $\mu' = \frac{\lambda(L+1)}{L}$.

We next prove (II) by observing that it is a variant of (I). Assume that we are given the *prior information* $r_0 = 1-\rho$. From Theorem 2, we have that $\kappa = \frac{\alpha}{\beta(1-\beta)} = \frac{1-k_0}{\beta}$. With some labour, using the mean queue length relation $L = (1-\rho)(1/k_0 - 1)/(1-\beta/k_0) = \rho(1+\rho y_s)/(1-\rho)$, it can be shown that $y_s = \frac{1-\kappa}{\kappa}$, from where equation (26) can be seen to reduce to equation (22). That is, $f(t)$ is identical to $B(t)$. A direct application of Theorem 2 yields

$$r_j = \rho\left[\frac{k_0}{\beta} - 1\right]\left[\frac{\beta}{k_0}\right]^j, \quad j \geq 1$$

$$= (1-\rho)\left[\frac{1-k_0}{\beta}\right]\left[\frac{\beta}{k_0}\right]^j, \quad j \geq 1$$

$$= (1-\rho)\left[\frac{1}{y_s+1}\right]\left[\rho^*\right]^j, \quad j \geq 1, \tag{27}$$

since $\kappa = \frac{1}{y_s+1}$, and using $\rho^* = \frac{\beta}{k_0}$. With $k_0 = 1-\beta\kappa$, and recalling that $\mu^* = \kappa\mu$, some simplification gives

$$\rho^* = \frac{\lambda(1+y_s)}{\mu^* + y_s(\mu^* + \lambda)}$$

$$= \frac{\lambda(1+y_s)}{\kappa\mu + y_s(\kappa\mu + \lambda)} = \frac{\rho(1+y_s)}{1+\rho y_s}$$

which proves that (27) is precisely the distribution (7) given by the El-Affendi and Kouvatsos theorem. Conversely, starting with the distribution in (27), a direct application of Theorem 2 proves that the service time distribution must be tainted exponential, of the form (26). ■

5. Conclusions

We have shown that the principle of entropy maximisation in M/GI/1 queues is equivalent to assuming that the queueing system is an M/M/1 queue. In addition, it as shown how the parameter of the exponential service time distribution may be chosen so as to yield the same equilibrium queue length distribution as given via EM. In an attempt to use EM and preserve $p_0 = 1-\rho$, it is possible to use the notion of "prior information", as was done in [8]. In this case, p_1 turns out to be the first term of a geometric sequence $\{p_n\}_{n=1}^{\infty}$, and our characterisation result tells us that the service time distribution must still possess a strongly exponential component. The service time distribution in this case is a mixture, with one component that is exponential, and another that is an impulse function of unit area at the origin. Since service times of length zero cannot occur in practice, this type of service time distribution raises some interesting questions. If we assume that zero length service times cannot occur, then the service time distribution reduces to an exponential distribution. Hence, in either case, there is a strong presence of exponentiality in the entropy maximisation method.

References

1. E.T. Jaynes, "Information theory and statistical mechanics," Phys. Rev., **106**, pp. 620–630, 1957.
2. V.E. Benes, *Mathematical Theory of Connecting Networks and Telephone Traffic*, Academic Press Inc., N.Y., 1965.
3. A.E. Ferdinand, "A Statistical Mechanical Approach to Systems Analysis," *I.B.M. J. Res. & Development*, September 1970.
4. ________________, "An analysis of the machine interference model," *I.B.M. Systems Journal*, No. 2, 1971.
5. D.D. Kouvatsos, "Maximum Entropy Methods for General Queueing Networks," Research Report RCC34, University of Bradford, Bradford, UK, 1983.
6. Y. Bard, "Estimation of State Probabilities using the Maximum Entroy Principle," *IBM J. Res. Dev.*, 24, 5, pp. 563–569, 1980.
7. J.E. Shore, "Information Theoretic Approximations for M/G/1 and G/G/1 Queueing Systems," *Acta Inf.*, 17, pp. 43–61, 1982.
8. M. El-Affendi and D. Kouvatsos, "A Maximum Entropy Analysis of the M/G/1 and G/M/1 Queueing Systems at Equilibrium," *Acta Inf.*, 19, pp. 339–355, 1983.
9. S. Guiasu, "Maximum Entropy Condition in Queueing Theory," *Journal of the Operational Res. Soc.*, Vol. 37, No. 3, pp. 293–301, 1986.
10. J. Engels and M. Zijlstra, "A Characterization of the Gamma Distribution by the Negative Binomial Distribution," *Journal of Appl. Prob.*, 17, pp. 1138–1144, 1980.
11. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd edn., Wiley, N.Y., 1971.
12. V. Rego, "Characterisations of Equilibrium Queue Length Distributions in M/GI/1 Queues," CSD-TR-616, Purdue University, 1986.
13. P.W. Purdom and C.A. Brown, *The Analysis of Algorithms*, Holt, Rinehart & Winston, N.Y., 1985.
14 Marcel F. Neuts, "The Caudal Characteristic Curve of Queues," *Adv. Appl. Prob.*, **18**, pp. 221-254, 1986.
15 D. Gross and C. Harris, *Fundamentals of Queueing Theory*, John Wiley, N.Y., 1985.
16 W. Szpankowski, "Some remarks on uniformly bounded Markov chains: Stability analysis," CSD-TR-600, Purdue University, 1986.