Contributed paper

A SUBJECTIVE BAYESIAN APPROACH TO THE THEORY OF QUEUES II – INFERENCE AND INFORMATION IN M/M/1 QUEUES

Michael F. McGRATH

Office of the Secretary of Defense
Washington, DC 20301, U.S.A.

and

Nozer D. SINGPURWALLA

Department of Operations Research
and Department of Statistics
The George Washington University
Washington, DC 20052, U.S.A.

Received 21 November 1985
(Revised 25 August 1986)

Abstract

This is a sequel to Part I of "A Subjective Bayesian Approach to the Theory of Queues". The focus here is on inference and a use of Shannon's measure of information for assessing the amount of information conveyed by the various types of data from queues. The notation and terminology used here is established in Part I.

Keywords

Bayesian analysis, Shannon information, subjective probability, discrete DFR distributions, mixtures of exponentials.

6. Inference in queues; preliminaries

In this section we address issues pertaining to the revision of probability statements regarding the behavior of a queueing system based on observed data. As stated earlier, the literature on this topic is exclusively sample theoretic, with the initial work by Clark [3] and Cox [4] outlining a general strategy, and the seminal paper by Basawa and Prabhu [2] culminating in an impressive demonstration of a use of this strategy. To discuss the various facets of a subjective Bayesian approach to inference in queues, we need to introduce certain concepts, notation and terminology and invoke these in the context of the M/M/1 queue.

A key step in undertaking inference in queues is a specification of the likelihood function. Difficulties here often stem from a failure to recognize that all probability statements are always conditional, the conditioning to be done on *all* the background information that is available when the probability statement is made. The background information must necessarily include the circumstances and conditions under which new information, say $\mathscr{X}$, is to be obtained. This means that consideration must be given to $\mathscr{D}$, the design of the experiment. Since the likelihood function is derived from a probability statement, it is clear that a change in $\mathscr{D}$ may lead to a change in the likelihood, with possible implications for inference. Bearing the above in mind, the next important issue in inference pertains to a *set of instructions* or a *rule* which either explicitly or implicitly tells us the process by which new information $\mathscr{X}$ is extracted. One of the functions of such a rule is to tell us when to begin and terminate the information extraction procedure. In the context of queueing theory, these instructions might typically include, inter alia: t_0, the time at which observation on a queueing process starts; T, the duration of time over which the queueing process is continuously observed; $M(N)$, the number of arrivals (services) for which the queueing process is to be continuously observed; $t_0, t_1, \ldots, t_k$, the time epochs at which "snap-shots" of the queueing process are taken, etc. Those quantities which are specified in advance of the experiments become a part of $\mathscr{D}$, the design of the experiment. Those quantities which become known to us after the experiment is performed constitute new information $\mathscr{X}$, and may include things such as: $N(t)$, the number in the system at time $t \in [t_0, t_0 + T]$; $t_0 + T_{M(N)}$, the time until the Mth(Nth) arrival (service completion); $A_1, A_2, \ldots, A_M$, the times between arrivals of the M customers; $S_1, S_2, \ldots, S_N$, the service times, etc. Thus, all quantities which describe observation of a queueing process have membership either in $\mathscr{D}$ or in $\mathscr{X}$, but not both, and the choice between $\mathscr{D}$ and $\mathscr{X}$ can vary from experiment to experiment, on the same queueing process.

Of particular concern to us is whether certain members of $\mathscr{X}$ are "informative" or "non-informative" with respect to a parameter of interest. The elements of $\mathscr{D}$, being specified in advance of experimentation, are necessarily non-informative. The above are important, because members of $\mathscr{X}$ which are non-information will *not* require an introduction of their probabilities in the likelihood function. Contrast this to the situation involving a sample-theoretic analysis which would involve an accounting of non-informative members of $\mathscr{X}$ (e.g. a "stopping rule") via an elaborate and sometimes complicated development. Thus, for example, Cox [4] in the last paragraph of page 294 says ". . . confidence intervals and significance tests would require for a careful analysis a specification of the stopping rule". By way of clarification, suppose that $\mathscr{X}$ consists of two members w_1 and w_2; then w_1 is said to be *non-informative* with respect to an unknown parameter $\boldsymbol{\theta}$, if w_1 is independent of $\boldsymbol{\theta}$, given w_2 – otherwise it is *informative* for $\boldsymbol{\theta}$. If w_1 is non-informative, then its probability cancels out in a Bayesian analysis of the problem, making w_1 *irrelevant* for inference.

To illustrate the above notions, suppose that $\mathscr{E}$ denotes the narrative specification of the experiment; then $\mathscr{E}$ and $\mathscr{D}$ become a part of $\mathscr{H}$, the background information. Given $\mathscr{X}$, inference for a parameter $\boldsymbol{\theta}$ is prescribed by Bayes law as

$$p(\boldsymbol{\theta}|\mathscr{X},\mathscr{H}) \propto p(\mathscr{X}|\boldsymbol{\theta},\mathscr{H})\,p(\boldsymbol{\theta}|\mathscr{H}), \tag{6.1}$$

where $p(\mathscr{X}|\boldsymbol{\theta},\mathscr{H})$, viewed as a function of $\boldsymbol{\theta}$ given $\mathscr{X}$, is the *likelihood* of $\boldsymbol{\theta}$. Suppose that $\mathscr{X} = (w_1, w_2)$ so that we may write

$$p(\mathscr{X}|\boldsymbol{\theta},\mathscr{H}) = p(w_1|w_2,\boldsymbol{\theta},\mathscr{H})\,p(w_2|\boldsymbol{\theta},\mathscr{H}). \tag{6.2}$$

If w_1 is non-informative about $\boldsymbol{\theta}$ given w_2, then $p(w_1|w_2,\boldsymbol{\theta},\mathscr{H})$ is independent of $\boldsymbol{\theta}$ and is thus a constant with respect to $\boldsymbol{\theta}$. Such members can be eliminated from the likelihood since they cancel out as constants of proportionality in an application of Bayes Law. Since the elements of $\mathscr{D}$ are specified in advance of experimentation, they are necessarily independent of $\boldsymbol{\theta}$ (although they could be dependent on *our knowledge* of $\boldsymbol{\theta}$), and are thus non-informative about $\boldsymbol{\theta}$. It is important to note that the informative or non-informative nature of any variable is a statement of (conditional) independence, and this according to the subjective Bayesian paradigm is a *judgement* by an analyst.

7. Experimental designs for the M/M/1 queue

In this section we consider some experimental designs for the M/M/1 queue, write out the elements $\mathscr{D}$ and $\mathscr{X}$ for each design and discuss the informative or non-informative nature of the elements of $\mathscr{X}$. As before, we let $\boldsymbol{\theta} = (\lambda, \mu)$ and let $\mathscr{I}_0 = \pi(\boldsymbol{\theta}|w)$ denote our current knowledge about $\boldsymbol{\theta}$. We start off by considering the simplest of the possible sampling schemes suggested by Cox [4].

Case 1: Here the scheme is described as follows:

$\mathscr{E}_1$: observe an M/M/1 queueing system continuously for a duration of $t(\mathscr{I}_0)$ units of time, commencing at time t_0, and record the interarrival and the service times *only*. Assume that the initial system size $N(t_0)$ is unobservable. The notation $t(\mathscr{I}_0)$ indicates the fact that our choice of the duration of observation could be influenced by our current knowledge of $\boldsymbol{\theta}$. If this choice is not influenced by $\mathscr{I}_0$, then $t(\mathscr{I}_0) = t$.

$\mathscr{D}_1$: $\{t_0, t(\mathscr{I}_0)\}$.

$\mathscr{X}_1$: $\left\{\left(A_1, \ldots, A_M, A_0 = \left(t(\mathscr{I}_0) - \sum_{j=1}^{M} A_j\right), M\right); (S_1, \ldots, S_N, N)\right\}$

where the A_i's (B_i's) are the interarrival (service) times and $M(N)$ are the number of arrivals (service completions) that are observed in $[t_0, t_0 + t(\mathcal{I}_0))$. If there is a customer who is in-service at $t_0\ [t_0 + t(\mathcal{I}_0)]$, then we shall assume that service for this customer has begun [completed] at $t_0\ [t_0 + (\mathcal{I}_0)]$.

Let $A = (A_1, \ldots, A_M, A_0)$, $S = (S_1, \ldots, S_N)$; also, let $a_i(s_i)$, $a(s)$, and $m(n)$ be realizations of $A_i(S_i)$, $A(S)$ and $M(N)$, respectively. Then

$$p(\mathcal{X}_1 \mid \boldsymbol{\theta}, \mathcal{H}) = p(m, n, \boldsymbol{a}, \boldsymbol{s} \mid t_0, t(\mathcal{I}_0), \boldsymbol{\theta}),$$

the likelihood of $\boldsymbol{\theta}$, can be written as the product

$$p(m, n \mid \boldsymbol{a}, \boldsymbol{s}, t_0, t(\mathcal{I}_0), \boldsymbol{\theta})\, p(\boldsymbol{a}, \boldsymbol{s} \mid t_0, t(\mathcal{I}_0), \boldsymbol{\theta})\, p(t_0, t(\mathcal{I}_0) \mid \boldsymbol{\theta}). \tag{7.1}$$

Clearly, the first term is degenerate at 1; this means that given A and $\boldsymbol{S}$, M and N are non-informative for $\boldsymbol{\theta}$, and thus we do not have to include their probabilities in the likelihood function. Contrast this to what would happen in a sample-theoretic analysis (other than point estimation by the method of maximum likelihood), wherein an introduction of the probabilities of M and N (the stopping rule) would be necessary and would complicate our analysis. Since our choice of t_0 and $t(\mathcal{I}_0)$ is independent of $\boldsymbol{\theta}$ – $t(\mathcal{I}_0)$ being dependent on $\mathcal{I}_0$, our current information about $\boldsymbol{\theta}$, rather than $\boldsymbol{\theta}$ itself – the third term of (7.1) is a constant independent of $\boldsymbol{\theta}$. Thus, all that matters for inference about $\boldsymbol{\theta}$ is the second term of (7.1), $p(\boldsymbol{a}, \boldsymbol{s} \mid t_0, t(\mathcal{I}_0), \boldsymbol{\theta})$, and this can be written as

$$\left(\prod_{i=1}^{m} \lambda\, e^{-\lambda a_i}\right) \left(\prod_{j=1}^{n} \mu\, e^{-\mu s_j}\right). \tag{7.2}$$

Inference for $\boldsymbol{\theta}$ given $\mathcal{X}_1$, assuming the prior $\pi(\boldsymbol{\theta} \mid \boldsymbol{w})$, follows via an application of Bayes Law, eq. (6.1). When $\pi(\boldsymbol{\theta} \mid \boldsymbol{w})$ is given by (5.1) of Part I, $\pi(\boldsymbol{\theta} \mid \mathcal{X}_1, \boldsymbol{w})$, the posterior density at $\boldsymbol{\theta}$ given $\mathcal{X}_1$, takes the simple form

$$C_1 \left(\lambda^{\alpha_1 + m - 1}\, e^{-\lambda(\beta_1 + \sum_{i=0}^{m} a_i)}\right) C_2 \left(\mu^{\alpha_2 + n - 1}\, e^{-\mu(\beta_2 + \sum_{j=1}^{n} s_j)}\right), \tag{7.3}$$

where C_1 and C_2 are constants. The above posterior is a product of two Erlang densities, so the appropriate material in sect. 5 follows by recursion. The choice of a conjugate (Erlang) prior thus leads to a convenient mechanism for updating statements of uncertainty regarding $\boldsymbol{\theta}$ and predicting the future behavior of the queue. When $\pi(\boldsymbol{\theta} \mid w)$ is given by (5.3), the recursive mechanism does not hold and a numerical approach becomes necessary – the details are not given here.

We note that (7.3) is equivalent to that obtained by Armero [1], whose work was unknown to us when our work was undertaken.

Case 2: Here the scheme is described as follows:

$\mathscr{E}_2$: observe an M/M/1 queueing system continuously, from time t_0 until the arrival of the $m(\mathscr{I}_0)^{\text{th}}$ customer (or alternatively, until the completion of the $n(\mathscr{I}_0)^{\text{th}}$ service), and record the interarrival and the service times only. The argument $\mathscr{I}_0$ with m reflects the fact that our choice of m could be based on our current knowledge of $\boldsymbol{\theta}$.

$\mathscr{D}_2$: $\{t_0, m(\mathscr{I}_0)\}$

$\mathscr{X}_2$: $\{A_1, A_2, \ldots, A_{m(\mathscr{I}_0)}); (S_1, S_2, \ldots, S_N, N)\}$.

We note here that when we fix $m(\mathscr{I}_0)$, the number of service completions N in the time interval $[t_0, t_0 + \Sigma_{i=1}^{m(\mathscr{I}_0)} A_i)$ would be random; thus, N would be a member of $\mathscr{X}_2$. This fact appears to have been overlooked by Lilliefors [8], Harris [7], and Schruben and Kulkarni [10], who in a sample-theoretic analysis of case 2 have failed to take into consideration the sampling distribution of N. Since N is non-informative for $\boldsymbol{\theta}$, when $\boldsymbol{A}$ and $\boldsymbol{S}$ are given, a Bayesian analysis of this case does not require an introduction of the probability of N and thus simplifies the calculation. To see the above, we note that the likelihood $p(n, \boldsymbol{a}, \boldsymbol{s} \mid t_0, m(\mathscr{I}_0), \boldsymbol{\theta})$ can be written as the product

$$p(n \mid \boldsymbol{a}, \boldsymbol{s}, t_0, m(\mathscr{I}_0), \boldsymbol{\theta})\, p(\boldsymbol{a}, \boldsymbol{s} \mid t_0, m(\mathscr{I}_0), \boldsymbol{\theta})\, p(t_0, m(\mathscr{I}_0) \mid \boldsymbol{\theta}),$$

with the first term being degenerate at 1 and the third term being independent of $\boldsymbol{\theta}$. Thus, all that matters for inference is the second term, which can be written as

$$\prod_{i=1}^{m(\mathscr{I}_0)} \lambda\, \mathrm{e}^{-\lambda a_i} \prod_{j=1}^{n} \mu\, \mathrm{e}^{-\mu s_j} . \tag{7.4}$$

Equation (7.4) is similar to (7.3), and the material following (7.3) applies here too. A similar analysis would apply if $n(\mathscr{I}_0)$ were fixed so that M would be random. It is not possible to fix both $n(\mathscr{I}_0)$ and $m(\mathscr{I}_0)$. This case illustrates the fact that *non-informative stopping rules,* as exemplified here by N, play no role in Bayesian inference for $\boldsymbol{\theta}$, whereas they do require an explicit accounting in a sample-theoretic approach to inference in queues.

Case 3: The scheme here is the following:

$\mathscr{E}_3$: observe an M/M/1 queueing system continuously for a duration of $t(\mathscr{I}_0)$ units of time, commencing at time t_0, and record the inter-arrival and interservice times and also $N(t_0)$, the number in the system at t_0.

$\mathscr{D}_3$: $\{t_0, t(\mathscr{I}_0)\}$, and

$\mathscr{X}_3$: $\{N(t_0), \mathscr{X}_1\}$, where $\mathscr{X}_1$ is described under case 1.

Let $n(t_0)$ be a realization of $N(t_0)$. Then the likelihood of $\boldsymbol{\theta}$, $p(m, n, n(t_0), \boldsymbol{a}, s \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta})$ can be factored as

$$p(m, n \mid n(t_0), \boldsymbol{a}, s, t_0, t(\mathscr{I}_0), \boldsymbol{\theta})\ p(\boldsymbol{a}, s \mid n(t_0), t_0, t(\mathscr{I}_0), \boldsymbol{\theta})$$

$$p(n(t_0) \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta})\ p(t_0, t(\mathscr{I}_0) \mid \boldsymbol{\theta}). \tag{7.5}$$

Since $\boldsymbol{A}$ and $\boldsymbol{S}$ are independent of $n(t_0)$, the second term of the above product can be written as $p(\boldsymbol{a}, s \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta})$, making (7.1) and (7.5) differ only by the term $p(n(t_0) \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta})$. Thus, as far as the likelihood is concerned, all that matters regarding inference for $\boldsymbol{\theta}$ is the product

$$p(\boldsymbol{a}, s \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta})\ p(n(t_0) \mid t_0, t(\mathscr{I}_0), \boldsymbol{\theta}), \tag{7.6}$$

for which the first term is given by eq. (7.2). The second term of (7.6) is more elaborate, it being given by eq. (4.1) [4.1a], with t replaced by t_0, n replaced by $n(t_0)$, and assuming that $N(0) = i$, for the (M/M/1/∞) [M/M/1/K] queue.

It is clear from the above that an evaluation of the likelihood (7.6), and $\pi(\boldsymbol{\theta} \mid \boldsymbol{w}, \mathscr{D}_3)$, the posterior at $\boldsymbol{\theta}$, will have to be undertaken numerically, as was done in sect. 5. This is a straightforward, but cumbersome, task. Thus, at least for the case of the prior given by (5.1), the inclusion of additional information provided by $N(t_0)$ has complicated our Bayesian analysis; the same would also be true of a sample-theoretic analysis. Cox [4] in his analysis tries to work around the problem by defining a "conditional likelihood" which de facto reduces to making $n(t_0)$ an element of $\mathscr{D}_3$ or by assuming that the queueing system has attained a steady state so that $p(n(t_0) \mid \cdot, \boldsymbol{\theta})$ may be written $(\lambda/\mu)^{n(t_0)}(1 - (\lambda/\mu))$. Both of the above strategies are questionable, since making $n(t_0)$ a member of $\mathscr{D}_3$ implies "start observation on the queueing process when the system size is $n(t_0)$" – a dubious proposition since the $n(t_0)$ may never be attained in a particular situation, and assuming a steady state when λ and μ are unknown and inference about them is sought is conceptually untenable.

In view of the above difficulties posed by the quantity $N(t_0)$, a fundamental question regarding the amount of information conveyed by $N(t_0)$ for inference about $\boldsymbol{\theta}$ arises. This question has also been raised by Cox, who provides an answer using a sample-theoretic approach based on a large sample theory and an assumption of the steady state. In sect. 8, we shall formally address this question within the Bayesian paradigm by using Shannon's measure of information.

Case 4: In this final case the scheme is described as follows:

$\mathscr{E}_4$: observe an M/M/1 queueing system at $(\ell + 1)$ discrete epochs of time $t_0, t_0 + \delta_1, \ldots, t_0 + \delta_\ell$, where $\delta_1 < \delta_2 < \ldots < \delta_\ell$ and record $N(j)$, the number in the system at the jth epoch; the quantities δ_i, $i = 1, \ldots, \ell$, and ℓ could be a function of $\mathscr{I}_0$ – for convenience, the argument $\mathscr{I}_0$ will be suppressed.

$\mathscr{D}_4$: $\{t_0, \ell, \delta_1 < \delta_2 < \ldots < \delta_\ell\}$.

$\mathscr{X}_4$: $\{N(t_0), N(t_0 + \delta_1), \ldots, N(t_0 + \delta_\ell)\}$.

Let $n(\cdot)$ be a realization of $N(\cdot)$. Then the likelihood of $\boldsymbol{\theta}$, $p(n(t_0), \ldots, n(t_0 + \delta_\ell) | \mathscr{D}_4, \boldsymbol{\theta})$ can be factored as

$$p(n(t_0 + \delta_\ell) | n(t_0 + \delta_{\ell-1}), \ldots, n(t_0), \mathscr{D}_4, \boldsymbol{\theta})$$

$$p(n(t_0 + \delta_{\ell-1}) | n(t_0 + \delta_{\ell-2}), \ldots, n(t_0), \mathscr{D}_4, \boldsymbol{\theta}) \ldots p(n(t_0) | \mathscr{D}_4, \boldsymbol{\theta}). \quad (7.7)$$

Due to the Markov property of $N(t)$ for the M/M/1 queue, (7.7) can be written as

$$p(n(t_0 + \delta_\ell) | n(t_0 + \delta_{\ell-1}), \mathscr{D}_4, \boldsymbol{\theta})$$

$$p(n(t_0 + \delta_{\ell-1}) | n(t_0 + \delta_{\ell-2}), \mathscr{D}_4, \boldsymbol{\theta}) \ldots p(n(t_0) | \mathscr{D}_4, \boldsymbol{\theta}). \quad (7.8)$$

Assuming that $N(0) = i$, the last term in the above expression is given by eq. (4.1) [4.1a], with t replaced by t_0, for the (M/M/1/∞) [M/M/1/K] queue. The other terms can similarly be evaluated by successively replacing n by $n(t_0 + \delta_\ell)$, $n(t_0 + \delta_{\ell-1}), \ldots, n(t_0 + \delta_1)$ and $N(0)$ by $n(t_0 + \delta_{\ell-1}), n(t_0 + \delta_{\ell-2}), \ldots, n(t_0)$, *mutatis-mutandis.*

Once again, it is clear from the above that the likelihood (7.8) and the posterior at $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} | w, \mathscr{X}_4)$, will have to be evaluated numerically, a cumbersome though straightforward task. Cox [4] encounters a similar difficulty in considering a sample-theoretic approch to the above problem, but circumvents it by assuming a steady state and focusing on inference about $\rho = (\lambda / \mu)$ via a likelihood which involves a product of terms of the form $\rho^{n(\cdot)}(1 - \rho)$. This strategy, besides being conceptually untenable, is necessarily inefficient since it ignores the Markov property of $N(t)$.

8. The use of Shannon's measure of information for inference in queues

In this section, we describe a general procedure which can be used to assess the amount of information provided by a member or a collection of members of $\mathscr{X}$ with regard to inference for $\boldsymbol{\theta}$. We then illustrate a use of this procedure for addressing the question of how much additional information is provided by $\mathscr{X}_3$ as compared to $\mathscr{X}_1$; recall that $\mathscr{X}_3 = (\mathscr{X}_1, N(t_0))$. The idea here is that if the additional information provided by $N(t_0)$ is small in comparison to that provided by $\mathscr{X}_1$, then we may base our inference for $\boldsymbol{\theta}$ on $\mathscr{X}_1$ alone and ignore $N(t_0)$, whose incorporation calls for much computational effort.

We should state at the outset that the Shannon measure of information which will be described here calls for much computation. However, such computations have to be undertaken only once for each of the various experiments that we wish to compare. In what follows, we shall undertake the above computations for a limited situation of interest. As stated before, our aim here is expository; we therefore confine ourselves to outlining the general principles and point out the various possibilities.

(a) AN OVERVIEW OF A MEASURE OF INFORMATION

Suppose that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a parameter of interest and $\mathscr{H}$ is some background information. Suppose that $\boldsymbol{\Theta}$ is endowed with a σ-field of subsets, and let $p(\boldsymbol{\theta} | \mathscr{H})$ be a prior probability density at $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with respect to a measure denoted by $\mathrm{d}\boldsymbol{\theta}$. In what follows, we shall suppress $\mathscr{H}$. Then, the *amount of information with respect* to $\mathrm{d}\boldsymbol{\theta}$ is defined by Shannon (c.f. Lindley [9]) is

$$\mathscr{I}_0 = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} ;$$

for any $\boldsymbol{\theta}$ for which $p(\boldsymbol{\theta}) = 0$, we define $p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) = 0$.

Suppose that an experiment $\mathscr{E}$ is performed and that $\mathscr{E}$ results in data $x \in \mathscr{X}$. The space $\mathscr{X}$ has a σ-field $\mathscr{B}$ of subsets X. For every $\boldsymbol{\theta} \in X$, a probability measure on $\mathscr{B}$ is defined; each probability measure is described by a probability density fuction $p(x | \boldsymbol{\theta})$ such that the probability measure of a subset X is given by $\int_X p(x | \boldsymbol{\theta}) \mathrm{d}x$. The ordered quadruple $\mathscr{E} = \{\mathscr{X}, \mathscr{B}, \boldsymbol{\Theta}, \mathscr{P}\}$, where $\mathscr{P}$ is the set of $p(x \mid \boldsymbol{\theta})$, characterizes an experiment $\mathscr{E}$. After the experiment is performed and x observed, the posterior distribution of $\boldsymbol{\theta}$ is $p(\boldsymbol{\theta} \mid x)$; the amount of information with respect to $\mathrm{d}\boldsymbol{\theta}$ is

$$\mathscr{I}_1(x) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta} \mid x) \log p(\boldsymbol{\theta} \mid x)\, \mathrm{d}\boldsymbol{\theta} .$$

The *amount of information provided* by $\mathscr{E}$ with prior knowledge $p(\boldsymbol{\theta})$, when x is observed is

$$\mathscr{I}(\mathscr{E}, p(\boldsymbol{\theta}), x) = \mathscr{I}_1(x) - \mathscr{I}_0 .$$

The above quantity depends on x; thus, we must average it with respect to x according to the predictive density

$$p(x) = \int_{\boldsymbol{\Theta}} p(x \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} .$$

Thus, the *average amount of information* provided by experiment $\mathscr{E}$, with prior knowledge $p(\boldsymbol{\theta})$ is

$$\mathscr{I}(\mathscr{E}, p(\boldsymbol{\theta})) = E_x[\mathscr{I}_1(x) - \mathscr{I}_0] , \tag{8.1}$$

where $E_x(\cdot)$ denotes expectation with respect to $p(x)$. When the particular prior distribution does not have to be stressed, (8.1) may be written $\mathscr{I}(\mathscr{E})$.

Suppose that $\mathscr{E}$ comprises of two experiments $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$, where $\mathscr{E}^{(i)} = \{\mathscr{X}^{(i)}, \mathscr{B}_i, \boldsymbol{\Theta}, \mathscr{P}_i\}$, where $\mathscr{P}_i$ is the set $p(x_i \mid \boldsymbol{\theta})$, $i = 1, 2$. We also need to consider the experiment $\mathscr{E}^{(2)}(x_1) = \{\mathscr{X}^{(2)}, \mathscr{B}_2, \boldsymbol{\Theta}, \mathscr{P}_2(x_1)\}$, where $\mathscr{P}_2(x_1)$ is the set of densities $p(x_2 \mid \boldsymbol{\theta}, x_1)$.

Consider $\mathscr{I}(\mathscr{E}^{(2)}(x_1), p(\boldsymbol{\theta} \mid x_1))$, the average information (with respect to $p(x_2 \mid \boldsymbol{\theta}, x_1)$) provided by an observation on $\mathscr{X}^{(2)}$ after $\mathscr{E}^{(1)}$ has been performed and x_1 observed. The average of the above, over x_1, is the average information provided by $\mathscr{E}^{(2)}$ after $\mathscr{E}^{(1)}$ has been performed. We suppress $p(\boldsymbol{\theta} \mid x_1)$ and denote the above by $\mathscr{I}(\mathscr{E}^{(2)} \mid \mathscr{E}^{(1)})$. We say that two experiments, $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ with $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$ are *independent* if $p(x_1, x_2 \mid \boldsymbol{\theta}) = p(x_1 \mid \boldsymbol{\theta})\, p(x_2 \mid \boldsymbol{\theta}), \forall\, \boldsymbol{\theta} \in \boldsymbol{\Theta}$. The following results given by Lindley [9] are relevant to us:

(a) $\mathscr{I}(\mathscr{E}) \geqslant 0$ with equality iff $p(x \mid \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$.

(b) $\mathscr{I}(\mathscr{E}^{(1)}) + \mathscr{I}(\mathscr{E}^{(2)} \mid \mathscr{E}^{(1)}) = \mathscr{I}(\mathscr{E})$.

(c) If X_1 is sufficient for $\boldsymbol{\theta}$ in the sense of Neyman and Fisher, then $\mathscr{I}(\mathscr{E}^{(1)}) = \mathscr{I}(\mathscr{E})$.

(d) If $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ are independent, then

(i) $\mathscr{I}(\mathscr{E}^{(2)}|\mathscr{E}^{(1)}) \leq \mathscr{I}(\mathscr{E}^{(2)})$, and

(ii) $\mathscr{I}(\mathscr{E}^{(1)}) + \mathscr{I}(\mathscr{E}^{(2)}) \geq \mathscr{I}(\mathscr{E})$

with equality iff X_1 and X_2 are independent.

The result (a) above says that any experiment is informative on the average, whereas (c) establishes no loss in information if attention is confined to observation on a sufficient statistic. If $\mathscr{E}^{(1)} \equiv \mathscr{E}^{(2)}$, then part (i) of (d) says that an independent repeat of the same experiment is less informative, on the average, than the original experiment.

(b) THE INFORMATION PROVIDED BY EXPERIMENT $\mathscr{E}_1$

We first consider the experiment $\mathscr{E}_1$ described under case 1 of sect. 7, and let $T_b = \Sigma_{j=1}^{N} S_j$, the "busy time", and $T_e = t(\mathscr{I}_0) - T_b$. Note that $(M, t(\mathscr{I}_0))$, (N, T_b) and (M, N, T_b, T_e) are sufficient for λ, μ, and $\boldsymbol{\theta}$, respectively. Thus, by property (c) of our measure of information

$$\mathscr{I}\left\{\mathscr{E}_1, \pi(\boldsymbol{\theta}|w, \mathscr{X}_1\right\} = \mathscr{I}\left\{\mathscr{E}_1, \pi(\boldsymbol{\theta}|w), (M, N, T_b, T_e)\right\}.$$

To evaluate the above, we may decompose $\mathscr{E}_1$ as $\mathscr{E}_1^{(1)}$ and $\mathscr{E}_1^{(2)}$, where $\mathscr{E}_1^{(1)} = \{M, \mathscr{B}_1, \boldsymbol{\Theta}, \mathscr{P}_1\}$ and $\mathscr{E}_1^{(2)} = \{(N, T_b), \mathscr{B}_2, \boldsymbol{\Theta}, \mathscr{P}_2\}$, where $\mathscr{B}_1$ and $\mathscr{B}_2$ are the σ-fields of $(0, 1, 2, \ldots,)$ and (N, T_b), respectively, $\mathscr{P}_1$ is the set $p(m|\boldsymbol{\theta}) = p(m|\lambda, \mu) = p(m|\lambda)$ and $\mathscr{P}_2$ the set $p(n, t_b|\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and t_b is a realization of T_b. Since $\mathscr{E}_1^{(1)}$ and $\mathscr{E}_1^{(2)}$ are characterized by the same $\boldsymbol{\Theta}$ and since $p(m, n, t_b|\boldsymbol{\theta}) = p(m|n, t_b, \boldsymbol{\theta})\cdot p(n, t_b|\boldsymbol{\theta}) = p(m|\boldsymbol{\theta})\, p(n, t_b|\boldsymbol{\theta})$, writing in a *temporal order* of occurrence of the variables, the experiments $\mathscr{E}_1^{(1)}$ and $\mathscr{E}_1^{(2)}$ may be judged independent. However, the random variables M and (N, T_b) are not independent* and so we cannot use the result that $\mathscr{I}(\mathscr{E}_1^{(1)}) + \mathscr{I}(\mathscr{E}_1^{(2)}) = \mathscr{I}(\mathscr{E})$. Rather, in our case $\mathscr{I}(\mathscr{E}_1^{(1)}) + \mathscr{I}(\mathscr{E}_1^{(2)}|\mathscr{E}_1^{(1)}) = \mathscr{I}(\mathscr{E})$.

We shall first consider the evaluation of $\mathscr{I}(\mathscr{E}_1^{(1)}, \pi(\lambda|w))$, where $\pi(\lambda|w)$ is an Erlang density with parameters α_1 and β_1. Thus

$$\mathscr{I}_0 = \int_0^\infty \frac{\beta_1^{\alpha_1}\, \lambda^{\alpha_1 - 1}\, e^{-\beta_1 \lambda}}{\Gamma(\alpha_1)} \log \frac{\beta_1^{\alpha_1}\, \lambda^{\alpha_1 - 1}\, e^{-\beta_1 \lambda}}{\Gamma(\alpha_1)}\, d\lambda$$

$$= \log\left(\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\right) + (\alpha_1 - 1)(\psi(\alpha_1) - \log \beta_1) - \alpha_1\,, \tag{8.2}$$

*To verify this, we extend the conversation to include $N(t_0)$ and note that $N \leq N(t_0) + M$.

where $\psi(\alpha_1)$ is Euler's psi function (Gradshteyn and Ryzhik [5], p. 943); $\psi(\alpha_1)$ can be evaluated recursively using the relationship $\psi(n+1) = \psi(n) + (1/n)$, with $\psi(1) = -0.57721\ldots$, the Euler constant.

After performing $\mathscr{E}_1^{(1)}$ and observing $M = m$, the posterior density at λ, given α_1, β_1, $t(\mathscr{I}_0)$ and m is also a gamma with shape parameter $\alpha_1 + m$ and scale parameter $\beta_1 + t(\mathscr{I}_0)$. Thus, $\mathscr{I}_1(m)$ is (8.2) with α_1 replaced by $\alpha_1 + m$, and β_1 replaced by $\beta_1 + t(\mathscr{I}_0)$, and

$$\mathscr{I}(\mathscr{E}_1^{(1)}) = E_m[\mathscr{I}_1(m) - \mathscr{I}_0]$$

$$= \sum_{m=0}^{\infty} [\mathscr{I}_1(m) - \mathscr{I}_0]\, f(m\,|\,\mathbf{w}),$$

where

$$f(m\,|\,\mathbf{w}) = \int_0^{\infty} f(m\,|\,\lambda)\,\pi_{\mathrm{I}}(\lambda\,|\,\mathbf{w})\,\mathrm{d}\lambda$$

$$= \binom{\alpha_1 + m - 1}{m} \left(\frac{\beta_1}{\beta_1 + t(\mathscr{I}_0)}\right) \left(1 - \frac{\beta_1}{\beta_1 + t(\mathscr{I}_0)}\right)^m, \tag{8.3}$$

a negative binomial density for $m = 0, 1, 2, \ldots$,; note that $f(m\,|\,\lambda)$ is a Poisson mass function with a rate λ and $\pi_{\mathrm{I}}(\lambda\,|\,\mathbf{w})$ is a gamma density. Thus, the average amount of information provided by experiment $\mathscr{E}_1^{(1)}$ with prior knowledge $\pi_{\mathrm{I}}(\lambda\,|\,\mathbf{w})$ is given by

$$\mathscr{I}(\mathscr{E}_1^{(1)}) = \sum_{m=0}^{\infty} \binom{\alpha_1 + m - 1}{m} \left(\frac{\beta_1}{\beta_1 + t(\mathscr{I}_0)}\right) \left(1 - \frac{\beta_1}{\beta_1 + t(\mathscr{I}_0)}\right)^m \tag{8.4}$$

$$\left[\log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1 + m)} + \log \frac{\beta_1 + t(\ _0)}{\beta_1} + (\alpha_1 + m - 1)\,\psi(\alpha_1 + m) - (\alpha_1 - 1)\,\psi(\alpha) - m\right].$$

We next consider the evaluation of $\mathscr{I}(\mathscr{E}_1^{(2)}|\mathscr{E}_1^{(1)}) = E_m[\mathscr{I}(\mathscr{E}_1^{(2)}(m), \pi(\boldsymbol{\theta}\,|\,m, \mathbf{w}))]$, where

$$\mathscr{I}(\mathscr{E}_1^{(2)}(m), \pi(\boldsymbol{\theta}\,|\,m, \mathbf{w})) = E_{t_b, n}[\mathscr{I}_1(t_b, n) - \mathscr{I}_1(m)].$$

Since $\pi(\mu\,|\,\mathbf{w})$ is an Erlang with parameters α_2 and β_2, $\pi(\mu\,|\,t_b, n, \mathbf{w})$ is Erlang with parameters $\alpha_2 + n$ and $\beta_2 + t_b$. Thus, $\mathscr{I}_1(t_b, n)$ is (8.2) with α_1 replaced by $\alpha_2 + n$ and β_1 replaced by $\beta_2 + t_b$. Thus,

$$\mathcal{I}(\mathcal{E}_1^{(2)}(m), \pi(\boldsymbol{\theta} \mid m, w)) = E_{t_b,n} \mathcal{I}_1(t_b, n) - \mathcal{I}_1(m),$$

where

$$E_{t_b,n} \mathcal{I}_1(t_b, n)$$

$$= E_{t_b,n} \left[\log \left(\frac{(\beta_2 + t_b)^{\alpha_2 + n}}{\Gamma(\alpha_2 + n)} \right) + (\alpha_2 + n - 1)(\psi(\alpha_2 + n) - \log(\beta_2 + t_b)) - (\alpha_2 + n) \right].$$

The computation of the above expectation is quite involved – see the development below; one possibility is to undertake it via a Monte Carlo simulation for different values of m, $m = 0, 1, 2, \ldots,$. Note that the above expectation is a function of m, and thus it is best that we denote this dependence by writing $E_{t_b,n} \mathcal{I}_1(t_b, n) = \mathcal{I}_1^\star(m)$. Once the expectation is computed, we may obtain

$$\mathcal{I}(\mathcal{E}_1^{(2)} \mid \mathcal{E}_1^{(1)}) = \sum_{m=0}^{\infty} [\mathcal{I}_1^\star(m) - \mathcal{I}_1(m)] \; f(m \mid w),$$

where $f(m \mid w)$ is given by (8.3). We then obtain

$$\mathcal{I}(\mathcal{E}_1) = \mathcal{I}(\mathcal{E}_1^{(1)}) + \mathcal{I}(\mathcal{E}_1^{(2)} \mid \mathcal{E}_1^{(1)}).$$

To appreciate the nature of the difficulties for computing $\mathcal{I}_1^\star(m)$, and also the fact that $E_{t_b,n} \mathcal{I}_1(t_b, n)$ is a function of m, we note that to compute the desired expectation we need to evaluate $p(t_n, n \mid t(\mathcal{I}_0))$. Extending the conversation to include $\boldsymbol{\theta}$, m, and $n(t_0)$, we have

$$p(t_b, n \mid t(\mathcal{I}_0))$$

$$= \int_{\boldsymbol{\Theta}} \sum_{m} \sum_{n(t_0)} p(t_b, n \mid \boldsymbol{\theta}, m, n(t_0)\, t(\mathcal{I}_0)\, p(\boldsymbol{\theta}, m, n(t_0) \mid w)\, \mathrm{d}\boldsymbol{\theta}$$

$$= \int_{\boldsymbol{\Theta}} \sum_{m} \sum_{n(t_0)} p(t_b \mid n, \boldsymbol{\theta}, m, n(t_0), t(\mathcal{I}_0))\, p(n \mid \boldsymbol{\theta}, m, n(t_0), t(\mathcal{I}_0))$$

$$p(m \mid \boldsymbol{\theta}, n(t_0), w)\, p(n(t_0) \mid \boldsymbol{\theta}, w)\, p(\boldsymbol{\theta} \mid w)\, \mathrm{d}\boldsymbol{\theta}$$

$$= \int_{\boldsymbol{\Theta}} \sum_{m} \sum_{n(t_0)} p(t_b \mid n, \boldsymbol{\theta})\, p(n \mid \boldsymbol{\theta}, m, n(t_0), t(\mathcal{I}_0))$$

$$p(m \mid \boldsymbol{\theta}, t(\mathcal{I}_0))\, p(n(t_0) \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid w)\, \mathrm{d}\boldsymbol{\theta}\,.$$

The first term in the above is an Erlang density with scale μ and shape n evaluated at t_b, whereas the third term is a Poisson mass function with a parameter $\lambda t(\mathcal{I}_0)$ evaluated at m. The fourth term is given by eq. (4.1) or [4.1a], whereas the fifth term is simply $\pi(\boldsymbol{\theta} \mid w)$. The second term $p(n \mid \boldsymbol{\theta}, m, n(t_0), t(\mathcal{I}_0))$ is difficult to evaluate; however, since $n = 0, 1, \ldots, m + n(t_0)$, the second term does indeed depend on m. It is for this reason that $\mathcal{I}_1^{\star}(m)$ has the index m.

It is clear from the above that a computation of $\mathcal{I}(\mathcal{E})$ is by no means a trivial exercise. All the same, it needs to be undertaken only once and for all. The same would also be true if we attempt to obtain $\mathcal{I}(\mathcal{E}_2)$ and $\mathcal{I}(\mathcal{E}_3)$, except that in the case of $\mathcal{I}(\mathcal{E}_3)$ we will have an additional variable $N(t_0)$ to worry about. The difficulty always arises when we try to evaluate $\mathcal{I}(\mathcal{E}_i^{(2)} \mid \mathcal{E}_i^{(1)})$, $i = 1, 2, 3$. As suggested before, one way out would be to perform a Monte Carlo simulation. This we have not done for now. However, since the motivation for undertaking the exercise is to evaluate the merits of observing $N(t_0)$ and accounting for it in terms of the additional information that is provided, we may focus attention on the experiment $\mathcal{E}_5$ described in the next section.

(c) EVALUATING THE ADDITIONAL INFORMATION PROVIDED BY $N(t_0)$

We shall now consider the following experiment:

$\mathcal{E}_5$: observe an M/M/1 queueing system continuously, from time t_0 until the arrival of the $m(\mathcal{I}_0) = m$th customer, and record $N(t_0)$ and the inter-arrival times. Assume that $N(0) = i$.

$\mathcal{D}_5$: $\{t_0, m, i\}$

$\mathcal{X}_5$: $\{N(t_0), A_1, A_2, \ldots, A_m\}$.

Let $T_a = \Sigma_{i=1}^{m} A_i$ and let t_a be a realization of T_a. Then T_a and $N(t_0)$ are jointly sufficient for $\boldsymbol{\theta}$. Thus, by property (c) of our measure of information

$$\mathcal{I}(\mathcal{E}_5, \pi(\boldsymbol{\theta} \mid w), \mathcal{X}_5) = \mathcal{I}(\mathcal{E}_5, \pi(\boldsymbol{\theta} \mid w), (N(t_0), t_a)). \tag{8.5}$$

To evaluate (6.14), we may decompose $\mathcal{E}_5$ into $\mathcal{E}_5^{(1)}$ and $\mathcal{E}_5^{(2)}$, where $\mathcal{E}_5^{(1)} = \{N(t_0), \mathcal{B}_1^{\star}, \boldsymbol{\Theta}, \mathcal{P}_1^{\star}\}$ and $\mathcal{E}_5^{(2)} = \{T_a, \mathcal{B}_2^{\star}, \boldsymbol{\Theta}, \mathcal{P}_2^{\star}\}$. $\mathcal{B}_1^{\star}$ and $\mathcal{B}_2^{\star}$ are the σ-fields of

$(0, 1, 2, \ldots,)$ and T_b, respectively. $\mathcal{P}_1^\star$ is the set $p(n(t_0)|\boldsymbol{\theta}, i)$, and $\mathcal{P}_2^\star$ the set $p(t_a|\boldsymbol{\theta}, m) = p(t_a|\lambda, m)$ for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. It is easy to verify that $\mathscr{E}_5^{(1)}$ and $\mathscr{E}_5^{(2)}$ are independent, but that $N(t_0)$ and T_a are not; thus, $\mathscr{I}(\mathscr{E}_5) = \mathscr{I}(\mathscr{E}_5^{(2)}) + \mathscr{I}(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)})$. In writing the last equality we have, for convenience, reversed the temporal order in which the experiments are performed.

Let us first consider the evaluation of $\mathscr{I}(\mathscr{E}_5^{(2)}, \pi(\lambda|w))$, where $\pi(\lambda\,|\,w)$ is an Erlang with parameters α_1 and β_1. Note that $\mathscr{E}_5^{(2)}$ provides us with information about λ alone and *not* about both λ and μ, it is for this reason that $\mathcal{P}_2^\star$ is the set $p(t_a|\lambda, m)$. It is now easy to verify that

$$\mathscr{I}(\mathscr{E}_5^{(2)}, \pi_{\mathrm{I}}(\lambda|w), t_a) = \mathscr{I}_1(t_a) - \mathscr{I}_0 = \log\left(\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1 + m)}\right) + \log\left(\frac{\beta_1 + t_a}{\beta_1}\right)$$

$$+ (\alpha_1 + m - 1)\,\psi(\alpha_1 + m) - (\alpha_1 - 1)\,\psi(\alpha_1) - m.$$

To obtain $\mathscr{I}(\mathscr{E}_5^{(2)})$, we need to evaluate

$$E_{t_a}[\mathscr{I}_{1(t_a)} - \mathscr{I}_0] = \int_0^\infty p(t_a)\,[\mathscr{I}_1(t_a) - \mathscr{I}_0]\,\mathrm{d}t_a,$$

where

$$p(t_a) = \int_0^\infty p(t_a|\lambda, m)\,\pi(\lambda|w)\,\mathrm{d}\lambda = \frac{\Gamma(\alpha_1 + m)\,\beta_1^{\alpha_1}}{\Gamma(\alpha_1)\,\Gamma(m)}\left(\frac{t_a^{m-1}}{(\beta_1 + t_a)^{\alpha_1 + m}}\right), \quad (8.6)$$

since $p(t_a|\lambda, m)$ is again an Erlang density with parameters λ and m. Some further straightforward calculations lead us to the result that

$$\mathscr{I}(\mathscr{E}_5^{(2)}) = \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1 + m)} + (\alpha_1 + m)\,\psi(\alpha_1 + m) - \alpha_1\,\psi(\alpha_1) - m. \quad (8.7)$$

It can be verified that $\mathscr{I}(\mathscr{E}_5^{(2)})$ is concave and increasing in m. This behavior supports the notion of *decreasing marginal utility* of observing an additional inter-arrival time.

We next consider the evaluation of $\mathscr{I}(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)})$. We first note that $\mathscr{E}_5^{(1)}$ provides us with information about *both* λ and μ, and therefore

$$\mathscr{I}(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)}) = E_{t_a}[\mathscr{I}(\mathscr{E}_5^{(1)}(t_a), \pi(\boldsymbol{\theta}|t_a, w))],$$

where

$$\mathcal{I}(\mathscr{E}_5^{(1)}(t_a), \pi(\boldsymbol{\theta}|t_a, w)) = E_{n(t_0)}[\mathcal{I}_1(n(t_0)) - \mathcal{I}_1(t_a)]$$

$$= E_{n(t_0)}\ \mathcal{I}_1(n(t_0)) - \mathcal{I}_1(t_a)).$$

To evaluate $\mathcal{I}_1(t_a)$, we first observe that upon observing t_a the posterior density at $\boldsymbol{\theta}$ given $\alpha_1, \alpha_2, \beta_1, \beta_2, m$ and t_a can be written (assuming (5.1) as the prior) as the product of two Erlang densities, one with parameters $\alpha = \alpha_1 + m$ and $\beta = \beta_1 + t_a$, and the other with parameters α_2 and β_2. Thus

$$\mathcal{I}_1(t_a) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}|t_a, w) \log \pi(\boldsymbol{\theta}|t_a, w)\, \mathrm{d}\boldsymbol{\theta}, \quad \text{or}$$

$$\mathcal{I}_1(t_a) = \left[\log\left(\frac{(\beta_1 + t_a)^{\alpha_1 + m}}{\Gamma(\alpha_1 + m)}\right) + (\alpha_1 + m - 1)(\psi(\alpha_1 + m) - \log(\beta_1 + t_a)) - \alpha_1 - m\right]$$

$$+ \left\{\log \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} + (\alpha_2 - 1)(\psi(\alpha_2) - \log \beta_2) - \alpha_2\right\}.$$

Note that the expression in the braces pertains to the amount of information with respect to $\mathrm{d}\mu$ that we have prior to accounting for $\mathscr{E}_5^{(1)}$, whereas the expression in the square brackets pertains to the amount of information with respect to $\mathrm{d}\lambda$ that we have prior to accounting for $\mathscr{E}_5^{(1)}$ but after accounting for $\mathscr{E}_5^{(2)}$. Recall that $\mathscr{E}_5^{(2)}$ does not give us any new information about μ. For future reference, let $\mathcal{I}_1^{\star}(t_a)$ denote $\mathcal{I}_1(t_a)$, given above, without the terms in the braces. That is

$$\mathcal{I}_1^{\star}(t_a) = \log\left(\frac{(\beta_1 + t_a)^{\alpha_1 + m}}{\Gamma(\alpha_1 + m)}\right) + (\alpha_1 + m - 1)(\psi(\alpha_1 + m) - \log(\beta + t_a)) - \alpha_1 - m.$$

We next consider the evaluation of

$$\mathcal{I}_1(n(t_0)) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}|n(t_0), t_a, i, m, w) \log \pi(\boldsymbol{\theta}|n(t_0), t_a, i, m, w)\, \mathrm{d}\boldsymbol{\theta}, \tag{8.8}$$

where $\pi(\boldsymbol{\theta}|n(t_0), t_a, i, m, w)$ is the posterior density at $\boldsymbol{\theta}$ given i, m, and w, and upon observing $n(t_0)$ and t_a. By Bayes rule, we note that

$$\pi(\boldsymbol{\theta}|n(t_0), t_a, i, m, w) \propto p(n(t_0), t_a|\boldsymbol{\theta}, i, m, w)\, p(\boldsymbol{\theta}|i, n, w)$$

$$= p(t_a|n(t_0), \boldsymbol{\theta}, i, m, w)\, p(n(t_0)|\boldsymbol{\theta}, i, m, w)\, p(\boldsymbol{\theta}|i, m, w)$$

$$= p(t_a|\boldsymbol{\theta}, m)\, p(n(t_0)|\boldsymbol{\theta}, i, w)\, p(\boldsymbol{\theta}|w)$$

$$= p(t_a|\lambda, m)\, (p_{i,n(t_0)}(t_0)|(\lambda, \mu))\, \pi(\boldsymbol{\theta}|w)\,. \tag{8.9}$$

To evaluate the above, we note that the first term is an Erlang density at t_a with scale λ and shape m, whereas the second term is given by (4.1) for the M/M/1/∞ queue and by [4.1a] for the M/M/1/K queue; the last term is of course given by (5.1). In view of the above, it appears that the evaluation of $\mathscr{I}_1(n(t_0))$ is best undertaken numerically.

Should we desire to focus attention on λ alone, then some simplification in the above-mentioned numerical exercise will result, since we now need to obtain

$$\mathscr{I}_1^\star(n(t_0)) = \int_0^\infty \pi(\lambda|n(t_0), t_a, i, m, w) \log \pi(\lambda|n(t_0), t_a, i, m)\, \mathrm{d}\lambda\,,$$

where

$$\pi(\lambda|\cdot) = \int_0^\infty \pi(\boldsymbol{\theta}|\cdot)\, \mathrm{d}\mu\,.$$

To obtain $\mathscr{I}(\mathscr{E}_5^{(1)}(t_a), \pi(\boldsymbol{\theta}|t_a, w))$, we need to compute $E_{n(t_0)}\, \mathscr{I}_1(n(t_0))$ where the expectation is to be taken with respect to the distribution given by $p_{i,n(t_0)}(t_0)$, where

$$p_{i,n(t_0)}(t_0) = \int_{\boldsymbol{\theta}} (p_{i,n(t_0)}(t_0)|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}|t_a, m, w)\, \mathrm{d}\boldsymbol{\theta}\,. \tag{8.10}$$

By Bayes rule, $\pi(\boldsymbol{\theta}|t_a, m, w) \propto p(t_a|m, \boldsymbol{\theta}, w)\, \pi(\boldsymbol{\theta}|w)$, where $p(t_a|m, \boldsymbol{\theta}, w)$ is the Erlang density at t_a with scale λ and shape m. By an analogous argument, we can also compute $E_{n(t_0)}\, \mathscr{I}_1^\star(n(t_0))$.

Our next step is the computation of $\mathscr{I}(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)}) = E_{t_a}[E_{n(t_0)}\, \mathscr{I}_1(n(t_0)) - \mathscr{I}_1(t_a)]$, where the expectation with respect to t_a is to be taken with respect to the distribution given by (8.6). Alternatively, should we wish to focus on the parameter λ only, we would compute $\mathscr{I}^\star(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)}) = E_{t_a}[E_{n(t_0)}\, \mathscr{I}_1^\star(n(t_0)) - \mathscr{I}_1^\star(t_a)]$. Once we obtain the above, we may compute $\mathscr{I}(\mathscr{E}_5) = \mathscr{I}(\mathscr{E}_5^{(2)}) + \mathscr{I}(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)})$ or $\mathscr{I}^\star(\mathscr{E}_5) = \mathscr{I}(\mathscr{E}_5^{(2)}) + \mathscr{I}^\star(\mathscr{E}_5^{(1)}|\mathscr{E}_5^{(2)})$.

To assess the amount of additional information provided by $N(t_0)$ as compared to T_a alone, we need to compare $\mathcal{J}(\mathscr{E}_5)$ versus $\mathcal{J}(\mathscr{E}_5^{(2)})$ or $\mathcal{J}^\star(\mathscr{E}_5)$ versus $\mathcal{J}(\mathscr{E}_5^{(2)})$ should we want to simplify matters by focusing on λ alone. In fig. 8.1, we show plots

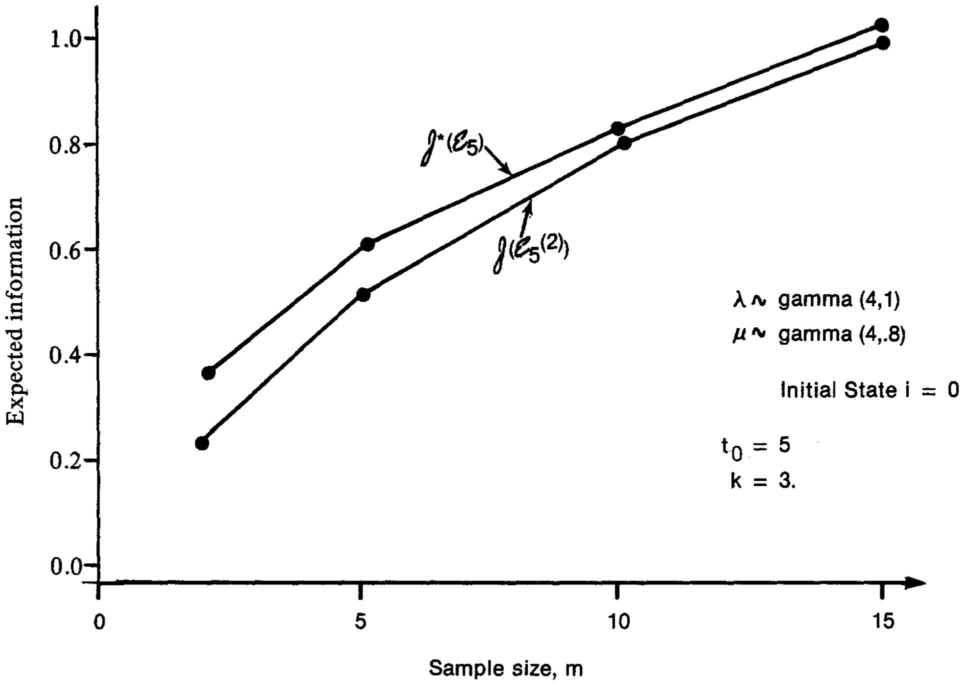

Fig. 8.1. A plot of the expected information versus the specified sample size m yielded by two experiments on an M/M/1/K queue.

of $\mathcal{J}^\star(\mathscr{E}_5)$ and $\mathcal{J}(\mathscr{E}_5^{(2)})$ versus m for the case $i = 0$, $t_0 = 5$, $K = 3$, $\alpha_1 = 4$, $\beta_1 = 1$, $\alpha_2 = 4$, and $\beta_2 = 0.8$. We note that as m increases, $\mathcal{J}^\star(\mathscr{E}_5)$ and $\mathcal{J}(\mathscr{E}_5^{(2)})$ tend to converge, implying that the additional information provided by $N(t_0)$ decreases in m, the number of observed interarrival times. Thus, for example, when $m > 10$, the relative loss in information with respect to dλ, $(\mathcal{J}^\star(\mathscr{E}_5) - \mathcal{J}(\mathscr{E}_5^{(2)}))/(\mathcal{J}^\star(\mathscr{E}_5)) = \mathcal{J}^\star(\mathscr{E}_5^{(1)} | \mathscr{E}_5^{(2)})/\mathcal{J}^\star(\mathscr{E}_5)$, when $N(t_0)$ is ignored is less than 3%. However, since $N(t_0)$ provides information about both λ and μ, the relative loss in information with respect to d$\boldsymbol{\theta}$ when $N(t_0)$ is ignored could be significant.

9. Extensions and conclusions

We have dealt with very simple queueing systems, the M/M/1/∞ and M/M/1/K, chosen both for ease of exposition and for the fact that closed form results are available for the transient and the steady-state measures of performance. Clearly, the subjective Bayesian approach is not limited to these simple models; any of the available models of classical queueing theory which are conditioned on $\boldsymbol{\theta}$ can be completed by averaging out the uncertainty with respect to the unknown parameters. A fertile area for future research is to use the approach of this paper for situations involving phase-type distributions and other more general situations such as multiserver queues, bulk service and arrival processes, queues with priority networks of queues, etc. (See Gross and Harris [6] for a compendium of the results available for such situations.) For most choices of prior distributions on $\boldsymbol{\theta}$, results can be obtained in a straightforward, albeit numerical, manner. For some of the classical Markovian queueing models (e.g. $M/M/\infty$), the choice of Erlang priors for the service and arrival parameters may yield results in analytical form for various measures of performance. It is not our intent to pursue such issues and details here.

In applications of classical queueing theory, a commonly used technique is to approximate a specified non-exponential input or service time distribution with a phase-type distribution such as Erlang or hyperexponential, and thus work with a Markovian model. From a subjective Bayesian standpoint, this is not an approximation at all – there is no "real" distribution to approximate. Provided that the shape and scale of the phase-type distribution (conditioned on $\boldsymbol{\theta}$) adequately reflects one's prior belief about the interarrival and service times, its use in the queueing model is a completely legitimate procedure. Of course, when the uncertainty regarding $\boldsymbol{\theta}$ is averaged out, the resulting distribution will, in general, no longer be phase type.

A prospect for further research would be the development of an interactive computer graphic technique to assist an analyst in describing his or her uncertainty about general interarrival and service times, representing these distributions as mixtures of exponentials and obtaining the predictive measures of performance of the queueing system.

As stated at the outset, the objective of this paper was to illustrate the subjective probability paradigm and its ramifications as a new way of looking at queueing systems. The motivation was on philosophical grounds, but the effects of the subjective viewpoint have pragmatic importance as well. We have developed a coherent approach to incorporating prior information (including expert opinion) about $\boldsymbol{\theta}$ into the model used to predict the behavior of the queueing system. We have used this coherent framework to incorporate observed data without dependence on assumptions of steady state or large samples. We have pointed out a use of the Shannon information measure as a means of comparing alternative experiments for observing queues. Finally, we have alluded to broad and (in our opinion) fertile areas for further research that can lead to substantial enrichment of queueing theory. We hope that exposure to the subjective probability paradigm might prompt queueing theorists to re-examine some long-held views and the pragmatic value of the results will be an adequate inducement to further consideration by practitioners of queueing theory.

Acknowledgements

We gratefully acknowledge the contribution of Professor I.V. Basawa, La Trobe University. His careful reading of Parts I and II and his helpful comments were the source of several important improvements in the final paper. This work was partially supported by the Army Research Office, Grant DAAG 29-84-K-0160 and the Office of Naval Research, Contract N00014-85-K-202, project NR 042-372.

References

[1] C. Armero, Bayesian analysis of M/M/1/∞, FIFO queues, *Bayesian Statistics, 2,* ed. J.M. Bernardo et al. (North-Holland, Amsterdam, 1985) pp. 613 – 618.

[2] I.V. Basawa and N.U. Prabhu, Estimation in single-server queues, Naval Research Logistics Quarterly 28(1981)475.

[3] A.B. Clark, Maximum likelihood estimates in a simple queue, Ann. Math. Statistics 28 (1957)1036.

[4] D.R. Cox, Some problems of statistical analysis connected with congestion, *Proc. Symp. on Congestion Theory,* ed. W.L. Smith and W.E. Wilkinson (University of North Carolina, 1964).

[5] I.S. Gradshteyn and Imryzhik, *Table of Integrals, Series and Products* (Academic Press, New York, 1980).

[6] D. Gross and C.M. Harris, *Fundamentals of Queueing Theory,* 2nd. ed. (Wiley, New York, 1985).

[7] C.M. Harris, Some new results in the statistical analysis of queues, *Mathematical Methods in Queueing Theory,* Proc. Conf. at Western Michigan University, ed. A.B. Clarke (Springer-Verlag, Berlin, 1974).

[8] H.W. Lilliefors, Some confidence intervals for queues, Oper. Res. 14(1966)723.

[9] D.V. Lindley, On a measure of the information provided by an experiment, Ann. Math. Statistics 27(1956)986.

[10] L. Schruben and R. Kulkarni, Some consequences of estimating parameters for the M/M/1 queue, Oper. Res. Lett. 1, 2(1982)75.