Bayesian analysis for the m/g/n queue using a phase type approximation

Yogesh Shukla, RK Shrivastava

Abstract
In this paper we study with Bayesian implication and forecasting for M/G/n queueing systems. The general service time density is approximated with a class of Erlang mixtures which are phase type distributions. Specified this phase type approximation, an explicit assessment of measures such as the stationary queue size, waiting time and busy period distributions can be obtained. Given arrival and service data, a Bayesian procedure based on reversible jump Markov Chain Monte Carlo methods is proposed to estimate system parameters and predictive distributions.

Keywords: Queues; Bayesian mixtures; reversible jump MCMC; phase type distributions, matrix geometric methods

Introduction
Bayesian analysis of queuing system seems to have been considered in early 1970’s; all derived statistical conclusions are obviously conditional on the assumed probability model. Unlike most other branches of mathematics, conventional methods of statistical inference suffer from the lack of an axiomatic basis; as a consequence, their proposed desiderata are often mutually incompatible, and the analysis of the same data may well lead to incompatible results when different, apparently intuitive procedures are tried: see Lindley (1972) [15] and by Jaynes (1976) [14], for many instructive examples. In marked contrast, the Bayesian approach to statistical inference is firmly based on axiomatic foundations which provide a unifying logical structure, and guarantee the mutual consistency of the methods proposed. Bayesian methods constitute a complete paradigm to statistical inference, a scientific revolution in Kuhn (1962) [16] sense. Bayesian statistics only require the mathematics of probability theory and the interpretation of probability which most closely corresponds to the standard use of this word in everyday language: it is no accident that some of the more important seminal books on Bayesian statistics, such as the works of de Laplace (1812) [13], Jeffreys (1939) [12] and de Finetti (1970) [11] or are actually entitled “Probability Theory”. The practical consequences of adopting the Bayesian paradigm are far reaching. Indeed, Bayesian methods (i) reduce statistical inference to problems in probability theory, thereby minimizing the need for completely new concepts, and (ii) serve to discriminate among conventional statistical techniques, by either providing a logical justification to some (and making explicit the conditions under which they are valid), or proving the logical inconsistency of others.

There has been much previous work on Bayesian density estimation using mixture models. See, for example, Diebort and Robert (1994) [4] MCMC methodology, Robert (1996) [6], have been developed for Bayesian analyses of mixture models. Recently, for variable dimension problems, often arising through model selection, Green (1995) [3] and Richardson and Green (1997) [2] introduced the reversible jump (RJ-MCMC) methods to analyze Normal mixtures. This type of algorithm was used by Rios et al. (1998) [9] for Exponential mixtures, Wiper et al. (2001) [10] for mixtures of Gamma distributions and Ausin et al. (2004) [1] for mixture of Erlang distributions.

More recently, in the context of mixtures of distributions, Stephens (2000a) rekindled interest in the use of continuous time birth-death methodology (BD-MCMC) for variable dimension problems. This type of methodology was used by Ausin and Wiper (2007a) [2] for mixtures of Erlang distributions. See Cappe et al., (2003) [3] for more details about RJ-MCMC and BD-
MCMC for mixture models. Mixture models allow a conveniently flexible family of distributions for estimating or approximating distributions which standard parametric family do not model appropriately, and provide a parametric alternative to non-parametric methods of density estimation. There has been much previous work on Bayesian density estimation using mixture models. See, for example, Diebolt and Robert (1994) [4] MCMC methodology, Robert (1996) [6], have been developed for Bayesian analyses of mixture models. Recently, for variable dimension problems, often arising through model selection, Green (1995) [5] and Richardson and Green (1997) [7] introduced the reversible jump (RJ-MCMC) methods to analyze Normal mixtures. This type of algorithm was used by Rios et al. (1998) [8] for Exponential mixtures, Wiper et al. (2001) [10] for mixtures of Gamma distributions and Ausin et al. (2004) [1] for mixture of Erlang distributions. More recently, in the context of mixtures of distributions, Stephens (2000a) rekindled interest in the use of continuous time birth-death methodology (BD-MCMC) for variable dimension problems. This type of methodology was used by Ausin and Wiper (2007a) [2] for mixtures of Erlang distributions. See Cappe et al. (2003) [3] for more details about RJ-MCMC and BD-MCMC for mixture models. Mixture models allow a conveniently flexible family of distributions for estimating or approximating distributions which standard parametric family do not model appropriately, and provide a parametric alternative to non-parametric methods of density estimation.

Methodology
An outline of our paper is as follows. Throughout we will assume the following FCFS queuing system. Let t be an inter arrival time; then t has an exponential distribution conditional on some unknown parameter λ, i.e.

\[ f(t|\lambda) = \lambda \exp(-\lambda t), \quad 0 < t < \infty \]

For service time S, we assume a mixture of k Erlang distribution with parameters W, μ, ν,

\[ f(S|k, W, \mu, \nu) = \sum_{i=1}^{k} w_i E_r(S|\nu_i, \mu_i) \]

Where

\[ E_r(S|\nu_i, \mu_i) = \frac{(\nu_i/\mu_i)^{\nu_i}}{\nu_i} S^{\nu_i-1} \exp \left( -\frac{\nu_i}{\mu_i} S \right) \]

And the mixture size, k and all other parameters are unknown.

Firstly we explain a simple experiment for observing arrival and service data. Prior distributions are defined for the unknown model parameters. We also mention the selection of simpler models for the service distribution.

We consider the problem of estimating the equilibrium characteristics of the queue. We briefly review the definitions and basic properties of phase type distribution. We illustrate how to obtain the stationary distribution of the queue size, the waiting time and the length of a busy period given the system parameters.

Given a sample realization of the Markov chain we have just defined and a sample of equal size from \( f(\lambda|t) \). we can estimate various quantities of interest. For example, given the sample data, we will often wish to assess whether or not the model is stable. The queue is stable if and only if the traffic intensity \( \rho \) is less than one. Thus we can estimate the probability of having a stable queue with

\[ P(\rho < 1|t, s) \approx \frac{1}{N} \#(\rho^n < 1) \]

Where

\[ \rho^n = \lambda^n \sum_{i=1}^{k} \omega_i^n \mu_i^n \]

And \( \{(k^{(1)}, w^{(1)}, \mu^{(1)}, \nu^{(1)}), \ldots, (k^{(n)}, w^{(n)}, \mu^{(n)}, \nu^{(n)})\} \) is a sample of size N obtained from the MCMC algorithm and \( \{\lambda^{(1)}, \ldots, \lambda^{(n)}\} \) is a sample of size N generated from the posterior distribution of \( \lambda \). A consistent estimator of the traffic intensity \( \rho \) is

\[ E(\rho|t, s) = E(\lambda|t) \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{k} \omega_i^n \mu_i^n \]

Where

\[ E(\lambda|t) = (a + m_p) \left( b + \sum_{j=1}^{m_q} t_j \right) \]

This provides a tool for service time model selection. For example, the posterior probability of having a single Erland distribution for the service Time distribution would be \( P(k=1|s) \). If this probability is big enough, we can model the system as an M/ER/1 queue and use a simpler algorithm to make inference about the system parameters, see Rios et al. (1998) [8].

Conclusion
We have developed a Bayesian analysis of M/G/1 system by modeling the general service time distribution using a mixture of Erlang distributions. We have constructed an MCMC algorithm making use of the reversible jump methodology and have combined this with matrix analytic method which has allowed us to make inference and predictions of various system quantities. We have illustrated our procedure with both simulated and real data. We have found some particular problem due to the discrete support of \( \nu \). Preservation of the second moments of the mixture in the reversible jump scheme is not possible.

References
2. Ausin MC, Wiper MP. Bayesian control of the number of servers in a GI/M/c queueing system, J Statist Planning and Inference. 2007a; 137:3043-3057.