# Bayesian inference and prediction in an M/G/1 with optional second service

Abdolreza Mohammadi, Mohammad Reza Salehi Rad

# Bayesian inference and prediction in an M/G/1 with optional second service

| | |
|---|---|
| Journal: | *Communications in Statistics - Simulation and Computation* |
| Manuscript ID: | LSSP-2011-0087.R2 |
| Manuscript Type: | Original Paper |
| Date Submitted by the Author: | 10-May-2011 |
| Complete List of Authors: | Mohammadi, Abdolreza; Johann Bernoulli Institute, Statistics<br>Salehi Rad, Mohammad Reza; Alame tabatabaii University, Statistics |
| Keywords: | M/G/1 queue, Optional service, Truncated Normal mixture, Birth-death MCMC |
| Abstract: | In this article, we exploit the Bayesian inference and prediction for an M/G/1 queuing system with optional second re-service. A service unit attends customers arriving following a Poisson process and demanding service according to a general distribution and some of customers need to re-service with probability "p". First, we introduce a mixture of truncated Normal distributions on to approximate the service and re-service time densities. Then, given observations of the system, we propose a Bayesian procedure based on birth-death MCMC methodology to estimate some performance measures. Finally, we apply the theories in practice by providing a numerical example. |
| | |

# *Bayesian inference and prediction in an M/G/1 with optional second service*

A. Mohammadi [1] , M. R. Salehi-Rad [2]

[1] Johann Bernoulli Institute, University of Groningen, Netherlands
E-mail: a.mohammadi@rug.nl

[2] Dept. of statistics, Allameh Tabataba`i University, Tehran, Iran
E-mail: salehirad@atu.ac.ir

**Abstract**

In this article, we exploit the Bayesian inference and prediction for an M/G/1 queuing system with optional second re-service. In this model, a service unit attends customers arriving following a Poisson process and demanding service according to a general distribution and some of customers need to re-service with probability "$p$" after taking the service. First, we introduce a mixture of truncated Normal distributions on interval $(-\infty,0)$ to approximate the service and re-service time densities of our queuing system. Then, given observations of the system, we propose a Bayesian procedure based on birth-death MCMC methodology to estimate some performance measures. Finally, we apply the theories in practice by providing a numerical example based on real data which have been obtained from a hospital.

**MSC:** 62F15, 60K25
**Keywords:** M/G/1 queue; Optional service; Truncated Normal mixture; Birth-death MCMC.

## 1. Introduction

The identification of a suitable queuing model to fitting the characteristics and description of the situation in hand is the first step that a queuing practitioner takes when analyzing a real life queuing problem. The following step usually aims at determination of performance measures. For instance, traffic intensity, mean system size, etc. These measures consequently are expressed as some function of parameters in the queuing system. Performance measures determination requires estimating of such parameters. Insufficient knowledge on numerical values of parameters presents a formidable hurdle to interpretation. How are such estimates obtained? Answering this question via the Bayesian approach constitutes the subject matter of this paper.

The model we have chosen to analyze is an M/G/1 queuing system with optional second service. While we are aware that some work in this direction has already been carried out, a few drawbacks have drawn our attention. Circumventing these drawbacks and suggesting improved methods was our main consideration. We have performed Bayesian inference to estimate the parameters of this queuing model. Bayesian inference is well suited for analyzing queues because we can evaluate posterior and predictive distributions for system performance measures taking different sources of uncertainties into account.

The majority of statistical papers on queuing systems are in the frequentist framework, with a renewed interest in Bayesian methods: M/M/1 queues are discussed by Armero and Bayarri (1994a,b), M/G/1 queues by Rios et al. (1998), M/G/1 queues using a phase-type approximation by Ausin et al. (2004), GI/G/1 queues by Ausin et al. (2007b) and GI/M/c queues by Ausin and Wiper (2007a). But in far applied queuing models, some customers might need to be re-serviced. For example, in a production line, some items might fail and require repair. In a hospital, some patients, after treatment, need extra services such as surgery. In these kinds of problems, we must re-service some items.

This paper constitutes Bayesian perspective for M/G/1 queuing systems with optional second service in which some customers must be re-serviced with probability $p$. Although, this queuing model has been studied by Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004) from classical queuing theory aspect; they have considered three alternatives for re-servicing in this queuing model and obtained the probability generating function (p.g.f.) of the steady-state system size at the moment of departure of the customer in the main queue, the mean busy period and the probability of the idle period. We will practically apply the Bayesian approach for one of these disciplines. Also, for this discipline we obtain the mean system size of the model in both queues via the p.g.f. of the system size.

The main contribution of this paper is that we will introduce a semi-parametric model for the service and re-service distributions based on a mixture of truncated Normal distributions on interval $(-\infty,0)$ that it will provide an alternative Bayesian approach for approximating the general distribution of service and re-service based on former papers.

There has been much previous work on Bayesian density estimation using mixture models. See, for example, Diebort and Robert (1994) MCMC methodology, Robert (1996), have been developed for Bayesian analyses of mixture models. Recently, for variable dimension problems, often arising through model selection, Green (1995) and Richardson and Green (1997) introduced the reversible jump (RJ-MCMC) methods to analyze Normal mixtures. This type of algorithm was used by Rios et al. (1998) for Exponential mixtures, Wiper et al. (2001) for mixtures of Gamma distributions and Ausin et al. (2004) 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) for mixtures of Erlang distributions. See Cappe et al. (2003) 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.

This paper is organized as follow. In section 2, we describe an M/G/1 queuing system with optional second service where we consider a semi-parametric approximation to the general distribution based on a mixture of truncated Normal distributions on interval $(-\infty,0)$. Also, we describe some results derived by Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004) which allow us to estimate the mean number of customers in the system, mean busy period and

probability of the idle period for our model.

In section 3, we define prior distributions and propose a birth-death MCMC technique to obtain a sample from the joint posterior distribution of the system parameters and the predictive service and re-service time distributions.

In section 4, we estimate the traffic intensity and the probability of holding the equilibrium condition stable. Also, we describe how we can estimate some performance measures by using the data generated from the birth-death MCMC algorithm.

In section 5, we illustrate our Bayesian methodology with the real data obtained from a hospital. Section 6 gives conclusions and possible extensions.

## 2. The queuing model

Throughout, we have considered an M/G/1 queuing model in a steady state, FCFS[1] discipline and independence between interarrival and service times in which some items failed and needs to be reserved with probability '$p$', independent of its service time or other factors (queuing system with two queues and one server.) . See Fig. 1.

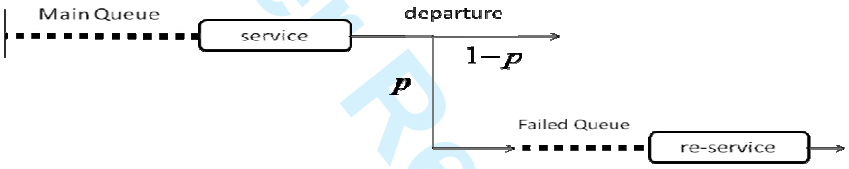


Fig.1. M/G/1 queuing model in which some items failed with probability '$p$', and required re-service

Moreover, we consider this model under a specific policy. We assume failed items are stockpiled in a failed queue (FQ) and re-serviced only after all customers in the main queue (MQ) are serviced. After completion of re-service of all items in FQ, the server returns to MQ if there are any customers waiting in MQ; otherwise, the server is idle. This queuing system was introduced by Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004) and considered from classical queuing theory.

In this model we suppose that a variable T is the interarrival time and has an exponential distribution with parameter $\lambda$. For service times, we suppose that service ($S$) and re-service ($\tilde{S}$) times are independent and have general distributions, denoted by $B_1(.)$ and $B_2(.)$ with means $\eta_1$, $\eta_2$ and variances $\delta_1, \delta_2$, respectively. For modeling the general time distributions of service and re-service, we alternatively introduce a semi-parametric model based on a mixture of truncated Normal distributions on interval $(-\infty, 0)$. Thus, if S is a typical service time, we have

$$B_1\left(s\left|\theta_1\right.\right)=\sum_{i=1}^{k_1}\pi_{1i}TN_{(0,\infty)}\left(s\left|\mu_{1i},\sigma_{1i}^2\right.\right), \qquad 0<s<\infty$$

1- First Come First Service

where $\underset{\sim}{\theta}_1=\left(k_1,\underset{\sim}{\pi}_1,\underset{\sim}{\mu}_1,\underset{\sim}{\sigma}_1^2\right)$, $k_1$ is the number of mixture components, $\underset{\sim}{\pi}_1=\left(\pi_{11},\pi_{12},...,\pi_{1k_1}\right)$ are weights and $TN_{(0,\infty)}\left(s\left|\mu_{1i},\sigma_{1i}^2\right.\right)$ represents the truncated Normal density function which has been truncated on interval $(-\infty,0)$, for $i=1,...,k_1$, that is,

$$TN_{(0,\infty)}\left(s\left|\mu_{1i},\sigma_{1i}^2\right.\right)=\frac{1}{\Phi\left(\mu_{1i}/\sigma_{1i}\right)\sqrt{2\pi\sigma_{1i}^2}}\exp\left(-\frac{1}{2\sigma_{1i}^2}(s-\mu_{1i})^2\right),\qquad s>0,\qquad(1)$$

in which $\Phi(.)$ is cumulative standard Normal distribution. Note that, in this mixture model all of the parameters are unknown specially the mixture size, $k_1$.

In the same way, if $\tilde{S}$ is a typical re-service time, we have,

$$B_2\left(\tilde{s}\left|\underset{\sim}{\theta}_2\right.\right)=\sum_{i=1}^{k_2}\pi_{2i}TN_{(0,\infty)}\left(\tilde{s}\left|\mu_{2i},\sigma_{2i}^2\right.\right),\qquad 0<\tilde{s}<\infty$$

where $\underset{\sim}{\theta}_2=\left(k_2,\underset{\sim}{\pi}_2,\underset{\sim}{\mu}_2,\underset{\sim}{\sigma}_2^2\right)$ and $TN_{(0,\infty)}\left(\tilde{s}\left|\mu_{2i},\sigma_{2i}^2\right.\right)$ is the same as defined for S.

Thus, we have a queuing system with two queues and one server with parameters $\underset{\sim}{\theta}=\left(\lambda,\underset{\sim}{\theta}_1,\underset{\sim}{\theta}_2,p\right)$, in which $\lambda$ is the parameter of interarrival times distribution and $p$ is the probability of the item being failed.

### 2.1. The mean system size

For our queuing model, under the steady state condition, suppose that $X_n$ is the number of customers remaining in MQ at the completion of the $n$ th customer's service time in MQ also $Y_n$ is the number of customers remaining in FQ at the completion of the $n$ th customer's service time in MQ. By using the p.g.f of the system size which has been obtained by Salehi-Rad and Mengersen (2002) we obtain the mean system size of the model in the next theorem. For a proof of this theorem see appendix.

**Theorem 2.1.** Expectation number of customers in MQ and FQ are as below

$$i)\qquad E(X_n)=\rho_1+\frac{\lambda^2\delta_1+\rho_1^2}{2(1-\rho_1)}+\frac{p}{2(1-\rho_1)}\left[\frac{\lambda^2\delta_2+\rho_2^2}{1-\rho_1}+p\rho_2^2\left(\frac{\lambda^2\delta_1+\rho_1(2-\rho_1)}{(1-\rho_1)^2}\right)\right]\pi_{0\bullet}\qquad(2)$$

$$ii)\qquad E\left(Y_n\right)=\frac{p\left[2\left(1-\rho_1\right)^2+\lambda^2\delta_1+\rho_1\left(1-\rho_1\right)\right]}{2\left(1-\rho_1\right)^2}\pi_{0\bullet}\qquad(3)$$

respectively. Where

$$\pi_{0\bullet}=(1-\rho_1)^2\left[p\rho_2+(1-\rho_1)G^*\left(0,p,\lambda\right)\right]^{-1}$$

$$G^*\left(0,p,\lambda\right)=\Psi\left(1-p+pB_2^*\left(\lambda\right)\right),$$

$$\Psi\left(u\right)=uB_1^*\left[\lambda\left(1-\Psi\left(u\right)\right)\right],$$

and $B_1^*(.)$ and $B_2^*(.)$ are the LSTs[2] of the distribution functions of the service and re-service times, respectively.

Proof. See Appendix.

Note that according to mixture of truncated Normal distributions for service and re-service times with some computations, the variance of service and re-service times are given by

$$\delta_j = \sum_{i=1}^{k_j} \pi_{ji}^2 \sigma_{ji}^2 \left[1 - \left(\Delta_{ji}\left(\Delta_{ji} + \frac{\mu_{ji}}{\sigma_{ji}}\right)\right)\right], \qquad j = 1,2$$

in which

$$\Delta_{ji} = \frac{\phi\left(\mu_{ji}/\sigma_{ji}\right)}{\Phi\left(\mu_{ji}/\sigma_{ji}\right)}$$

that $\phi(.)$ is a standard Normal density.

### 2.2. The mean busy period and probability of idle period of the system

To obtain the mean busy period for our queuing model, we have

E[busy period] = E[busy period in MQ] + $p$ E[busy period in FQ]

The first expressions is equal to $\eta_1/(1-\lambda\eta_1)$ and the second expression is $p\eta_2/(1-\rho_1)$. Thus

$$\mathrm{E}\left[\text{busy period}\right] = \frac{\eta_1 + \eta_2 p^2}{1-\lambda\eta_1} \tag{4}$$

Note that, according to mixture of truncated Normal distributions for service and re-service times with some computations, we have

$$\eta_j = \left[\sum_{i=1}^{k_j} \pi_{ji}\left(\mu_{ji} + \sigma_{ji}\frac{\phi\left(\mu_{ji}/\sigma_{ji}\right)}{\Phi\left(\mu_{ji}/\sigma_{ji}\right)}\right)\right], \qquad j = 1,2$$

For finding the probability of idle period, we have

$$\pi^* = \Pr\left(\text{idle period}\right) = \frac{1-\rho_1}{1+p^2\rho_2}. \tag{5}$$

For more details see Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004).

## 3. Bayesian inference

In this section, we develop Bayesian inference techniques for system parameters $\underset{\sim}{\theta} = \left(\lambda, \underset{\sim}{\theta_1}, \underset{\sim}{\theta_2}, p\right)$. We consider throughout the simple experiment of observing $n_a$ interarrival times $\underset{\sim}{t} = \left\{t_1,...,t_{n_a}\right\}$, $n_{s1}$ service times in MQ, as $\underset{\sim}{s_1} = \left\{s_{11}, s_{12},...,s_{1n_{s1}}\right\}$, $n_{s2}$ re-service

2- Laplace Stieltjes Transform

times in FQ, as $\underset{\sim}{s}_2=\left\{s_{21},s_{22},...,s_{2n_{s2}}\right\}$ and $n_p$ sample for parameter *p*, as $\underset{\sim}{x}=\left\{x_1,...,x_{n_p}\right\}$ in which $x_i=1$ if customer need re-service and $x_i=0$ if customer does not need re-service, for $i=1,...,n_p$. Given these observations, $\underset{\sim}{t}$, $\underset{\sim}{s}_1$, $\underset{\sim}{s}_2$, and $\underset{\sim}{x}$, the likelihood function of the queuing system under our policy is

$$L\left(\underset{\sim}{\theta}|\underset{\sim}{t},\underset{\sim}{s}_1,\underset{\sim}{s}_2,\underset{\sim}{x}\right)\propto L\left(\lambda|\underset{\sim}{t}\right)L\left(\underset{\sim}{\theta}_1|\underset{\sim}{s}_1\right)L\left(\underset{\sim}{\theta}_2|\underset{\sim}{s}_2\right)L\left(p|\underset{\sim}{x}\right)$$

in which

$$L\left(\lambda|\underset{\sim}{t}\right)=\lambda^{n_a}\exp\left(-\lambda\sum_{i=1}^{n_a}t_j\right),$$

$$L\left(\underset{\sim}{\theta}_1|\underset{\sim}{s}_1\right)=\prod_{j=1}^{n_{s1}}\left(\sum_{i=1}^{k_1}\pi_{1i}TN_{(0,\infty)}\left(s_{1j}\middle|\mu_{1i},\sigma_{1i}^2\right)\right),$$

$$L\left(\underset{\sim}{\theta}_2|\underset{\sim}{s}_2\right)=\prod_{j=1}^{n_{s2}}\left(\sum_{i=1}^{k_2}\pi_{2i}TN_{(0,\infty)}\left(s_{2j}\middle|\mu_{2i},\sigma_{2i}^2\right)\right),$$

and

$$L\left(p|\underset{\sim}{x}\right)=p^{\sum x_i}\left(1-p\right)^{n_p-\sum x_i}.$$

**Remark 3.1.** In this model we assume those interarrival times, service times, re-service times and the probability of re-service, are independent.

### 3.1. Prior specification and updating

Here, we assign prior distributions for the parameters of our queuing system. For the arrival rate, $\lambda$, we assume a gamma distribution, $\lambda\sim G\left(a,b\right)$. It is straightforward to show that, conditional on arrival data, the posterior distribution of $\lambda$ is also a gamma distribution as $G\left(a+n_a,b+\sum_{i=1}^{n_a}t_i\right)$, see, for example, Armero and Bayarri (1994b).

For the parameter *p*, we assume a Beta distribution, $P\sim Beta\left(\varepsilon,\zeta\right)$. It is easy to show that the posterior distribution of *p* is $Beta\left(\varepsilon+\sum_{i=1}^{n_p}x_i,\zeta+n_p-\sum_{i=1}^{n_p}x_i\right)$.

Since both of $B_1(.)$ and $B_2(.)$ are approximated with the mixture of truncated Normal distributions on interval $(-\infty,0)$, so the Bayesian model for service distribution parameters and for re-service distribution parameters are the same.

In order to make inference for service distribution parameters, $\underset{\sim}{\theta}_1$, as in Diebolt and Robert (1994), it is convenient to assume that each observation $S_{1i}$ arose from an unknown component $Z_{1i}$ of the mixture, where $z_{11},z_{12},...,z_{1n_{s1}}$ are realizations of independent and identically distributed discrete random variables $Z_{11},Z_{12},...,Z_{1n_{s1}}$ with the prior distributions as below

$$P\left(Z_{1i}=j\,|k_1,\pi_1\right)=\pi_{1j}\,,\quad \left(i=1,...,n_{s1};\,j=1,...,k_1\right)$$

Then, the conditional distribution of the service time, $S_{1i}$, given $Z_{1i}$, is

$$S_{1i}\left|Z_{1i}=j\right. \sim TN_{(0,\infty)}\left(s\left|\mu_{1j},\sigma_{1j}^2\right.\right),\qquad \left(i=1,...,n_{s1}\right).$$

Now, we can define a joint prior distribution on the mixture parameters of the service time, $\underset{\sim}{\theta}_1=\left(k_1,\underset{\sim}{\pi}_1,\underset{\sim}{\mu}_1,\underset{\sim}{\sigma}_1^2\right)$. We will assume that the joint prior distribution can be factorized as

$$f\left(\underset{\sim}{\pi}_1,\underset{\sim}{\mu}_1,\underset{\sim}{\sigma}_1^2\left|\underset{\sim}{z}_1,k_1\right.\right)\propto f\left(\underset{\sim}{z}_1\left|k_1,\underset{\sim}{\pi}_1\right.\right)f\left(\underset{\sim}{\pi}_1\left|k_1\right.\right)f\left(\underset{\sim}{\mu}_1\left|k_1\right.\right)f\left(\underset{\sim}{\sigma}_1^2\left|k_1\right.\right)$$

see Richardson and Green (1997). We now define proper prior distribution for $\underset{\sim}{\pi}_1,\underset{\sim}{\mu}_1,\underset{\sim}{\sigma}_1^2$.

First, following Stephens (2000a) we assume a truncated Poisson distribution for the mixture size, $k_1$, taking values on 1 to $k_{1\max}$, that is

$$P\left(k_1=k\right)\propto\frac{\gamma^k}{k!},\quad k=1,...,k_{\max} \tag{6}$$

We take $k_{\max}=40$ in order to penalize against over-fitting the data with a large number of components. We also define prior distributions for remaining parameters given that $k_1$, as below

$$\begin{aligned}
&\underset{\sim}{\pi}_1\left|k_1\right. \sim D\left(\phi_{11},...,\phi_{1k_1}\right)\\
&\mu_{1i}\left|k_1\right. \sim TN_{(0,\infty)}\left(\xi,\tau\right)\qquad for\quad i=1,...,k_1\\
&\sigma_{1i}^2\left|k_1\right. \sim IG(\alpha,\beta)\qquad for\quad i=1,...,k_1
\end{aligned}$$

where $D\left(\phi_{11},...,\phi_{1k_1}\right)$ denotes a Dirichlet distribution with parameters $\phi_{1r}>0$ for $r=1,...,k_1$ with density

$$f\left(\underset{\sim}{\pi}_1\left|k_1\right.\right)=\frac{\Gamma\left(\phi_{11}+\cdots+\phi_{1k_1}\right)}{\Gamma\left(\phi_{11}\right)\cdots\Gamma\left(\phi_{1k_1}\right)}\prod_{r=1}^{k_1}\pi_{1r}^{\phi_{1r}-1}$$

Typically, we might set $\phi_{1r}=1$, for all $r=1,...,k_1$, giving a uniform U(0,1) prior for the weights. $TN_{(0,\infty)}\left(\xi,\tau\right)$ denotes a truncated Normal on interval $(-\infty,0)$, see formula (1).

Finally, $IG\left(\alpha,\beta\right)$ denotes an inverted gamma distribution with density

$$f\left(\sigma_{1i}^2\left|k_1\right.\right)=\frac{\beta^\alpha}{\Gamma\left(\alpha\right)}\left(\sigma_{1i}^2\right)^{-(\alpha+1)}\exp\left(-\frac{\beta}{\sigma_{1i}^2}\right)$$

**Remark 3.2.** Note that, the mean of service time can not be negative so we suppose a truncated Normal on interval $(-\infty,0)$ for prior distribution of $\mu_{1i}$.

Note that our model is invariant to permutation of the labels $i=1,...,k_1$. For identifiability, it is important to adopt a unique labeling. Unless stated otherwise, we use that in which the $\mu_{1i}$ are increasing numerical order; thus the joint prior distribution of the parameters is $k_1!$ times the product of the individual truncated Normal and inverted gamma densities, restricted to the set

$\mu_{11} < \mu_{12} < \cdots < \mu_{1k_1}$, see Stephens (2000a,2000b).

Conditional on $k_1$ and given the service time data, the full conditional posterior distribution of the allocation variables is

$$P\left(Z_{1j}=i|\cdots\right) \propto \pi_{1i} TN_{(0,\infty)}\left(s_{1i}\middle|\mu_{1i},\sigma_{1i}^2\right) \ , i=1,...,k_1$$

here and later we use '$|\cdots$'to denote conditioning on all other variables.

Full conditional for the weights $\underset{\sim}{\pi}_1$ remains Dirichlet as

$$\underset{\sim}{\pi}_1|\cdots \sim D\left(\phi_{11}+n_{11},...,\phi_{1k_1}+n_{1k_1}\right)$$

where $n_{1i} = \#\left\{z_{1j}=i\right\}$. Also, full conditionals for $\left\{\mu_{1i}\right\}$ can be derived as below

$$\mu_{1i}|\cdots \propto \frac{1}{\Phi\left(\mu_{1i}/\sigma_{1i}\right)} TN_{(0,\infty)}\left(\frac{\frac{1}{\sigma_{1i}^2}\sum_{j:j=z_i} s_{1j}+\frac{\xi}{\tau}}{\frac{1}{\tau}+\frac{n_{1i}}{\sigma_{1i}^2}}, \frac{1}{\frac{1}{\tau}+\frac{n_{1i}}{\sigma_{1i}^2}}\right), \qquad \mu_{1i}>0$$

Note that, to preserve the ordering constraint on the $\left\{\mu_{1i}\right\}$, the full conditional is used only to generate a proposal and is accepted provided that the ordering is unchanged.

The full conditionals for $\left\{\sigma_{1i}^2\right\}$ are

$$\sigma_{1i}^2|\cdots \propto \frac{1}{\Phi\left(\mu_{1i}/\sigma_{1i}\right)} IG\left(\alpha+\frac{n_{1i}}{2}, \beta+\frac{1}{2}\sum_{j:j=z_i}\left(s_{1j}-\mu_{1i}\right)^2\right)$$

Now under our posterior distributions we can define a BD-MCMC algorithm which is considered in next section.

### 3.2. BD-MCMC algorithm

Here, we propose a birth-death MCMC algorithm for sampling from the full posterior distribution of service parameters, i.e. $\underset{\sim}{\theta}_1=\left(k_1,\underset{\sim}{\pi}_1,\underset{\sim}{\mu}_1,\underset{\sim}{\sigma}_1^2\right)$, for more details, including details of the construction of the BD-MCMC methodology, see Stephens (2000a) and Cappe (2003). The DBMCMC approach was introduced by Stephens (2000a) for Normal mixtures which is based on a general Markov birth-death process (BD). In this approach, the mixture size, $k_1$, changes such that births and deaths of the mixture components occur in continuous time with the stationary distribution of a joint posterior of the mixture parameters. In order to improve mixing, the BD process can be combined with a standard MCMC method where $k_1$ is kept fixed, as will be shown further on. This methodology has been used for an Erlang mixture model to approximate the service distribution by Ausin and Wiper (2007a).

In the BD process, births of mixture components occur at a constant rate which we might set equal to the parameter, $\gamma$, from the prior distribution of $k_1$ in (6). A birth increases the number

of components by one. The weight of a new component is generated from $Beta(1,k_1)$ and the remaining parameters are sampled from the prior distributions. The death rate of every mixture component is a likelihood ratio of the model with and without this component, given by

$$\omega_j = \prod_{r=1}^{n_{s1}} \left( \frac{\sum_{\substack{i=1 \\ i\neq j}}^{k_1} \pi_{1i} \Big/ (1-\pi_{1j}) TN_{(0,\infty)}\left(s_{1r} \middle| \mu_{1i}, \sigma_{1i}^2\right)}{\sum_{i=1}^{k_1} \pi_{1i} TN_{(0,\infty)}\left(s_{1r} \middle| \mu_{1i}, \sigma_{1i}^2\right)} \right), \qquad j = 1,...,k_1$$

Thus, death rates are very low if the corresponding component explains a lot of data and high if it does not. The total death rate, i.e. $\omega = \sum_j \omega_j$ , of the process at any time is the sum of the individual death rates. A death decreases the number of mixture components by one. The birth and death processes are independent Poisson processes, thus, the time of birth/death event is exponentially distributed with mean $1/(\omega+\gamma)$ and a birth or death occur with probabilities proportional to $\gamma$ and $\omega$, respectively.

The following algorithm defines a general Markov berth-death process with the stationary distribution $B_1\left(s \middle| \underset{\sim}{\theta}_1\right)$, based on Stephens (2000a). Also this algorithm used by Ausin and Wiper (2007a) for mixture of Erlang distributions.

**Algorithm 3.1.** Starting with initial values $k_1^{(0)}, \underset{\sim}{\pi}_1^{(0)}, \underset{\sim}{\mu}_1^{(0)}, \underset{\sim}{\sigma}_1^{2(0)}$, iterate the following steps:

**Birth-death process**

1. Run the birth-death process for a fixed time $t_0$,
   1.1. Start from $k_1^{(n)}, \underset{\sim}{\pi}_1^{(n)}, \underset{\sim}{\mu}_1^{(n)}, \underset{\sim}{\sigma}_1^{2(n)}$,
   1.2. Compute the death rates,
   1.3. Simulate the exponential time to next jump,
   1.4. Simulate the type of jump (birth or death),
   1.5. Modify the mixture components and
   1.6. If the run time is less than $t_0$ go to 1.2,

**MCMC algorithm conditional on** $k_1$

2. Generate $\underset{\sim}{z}_1^{(n+1)} \sim \underset{\sim}{z}_1 \middle| \underset{\sim}{s}_1, k_1^{(n+1)}, \underset{\sim}{\pi}_1^{(n)}, \underset{\sim}{\mu}_1^{(n)}, \underset{\sim}{\sigma}_1^{2(n)}$,
3. Generate $\underset{\sim}{\pi}_1^{(n+1)} \sim \underset{\sim}{\pi}_1 \middle| \underset{\sim}{s}_1, k_1^{(n+1)}, \underset{\sim}{z}_1^{(n+1)}$,
4. for $r = 1,...,k_1^{(n+1)}$,
   4.1. Generate $\mu_{1r}^{(n+1)}$ using a Metropolis step.
   4.2. Generate $\sigma_{1r}^{2(n+1)}$ using Metropolis step.
5. Set $n = n+1$ and go to 1.

Note that the steps 2 to 3 are standard Gibbs sampling, whereby the model parameters are updated conditional on the mixture size, $k_1$. The only complicated is step 4, where we introduce

a Metropolis Hasting method. To do this, for step 4.1 we generate candidate values for $\mu_{1r}^{(n+1)}$ from a proposal distribution as below

$$TN_{(0,\infty)}\left(\frac{\frac{1}{\sigma_{1i}^2}\sum_{j:j=z_i} s_{1j}+\frac{\xi}{\tau}}{\frac{1}{\tau}+\frac{n_{1i}}{\sigma_{1i}^2}},\frac{1}{\frac{1}{\tau}+\frac{n_{1i}}{\sigma_{1i}^2}}\right).$$

Also, for step 4.2 we generate candidate values for $\sigma_{1r}^{2\,(n+1)}$ from a proposal distribution as below

$$IG\left(\alpha+\frac{n_{1i}}{2},\beta+\frac{1}{2}\sum_{j:j=z_i}\left(s_{1j}-\mu_{1i}\right)^2\right).$$

This algorithm produces a sample from the joint posterior distribution. Given suitable regularity conditions, see Tierney (1995), quantities of interest may be consistently estimated by sample path averages. Thus, given the MCMC output of size N, we can estimate the mixture size of service and re-service distribution, $k_1$ and $k_2$. We do only for $k_1$ and $k_2$ is the same. So, we estimate the marginal posterior distribution by

$$P\left(k_1=k\left|\underset{\sim}{t},\underset{\sim}{x},\underset{\sim}{s}_1,\underset{\sim}{s}_2\right.\right)\approx\frac{1}{N}\#\left\{n:k_1^{(n)}=k\right\} \tag{7}$$

We can determine the number of phases of service distribution with this probability.

We can also perform inference for the predictive density of the service time distribution by

$$B_1\left(s\left|\underset{\sim}{s}_1\right.\right)\approx\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{k_1^{(n)}}\pi_{1i}^{(n)}TN_{(0,\infty)}\left(s\left|\mu_{1i}^{(n)},\sigma_{1i}^{2(n)}\right.\right) \tag{8}$$

The estimation of mixture distribution parameters of re-service times, i.e. $\underset{\sim}{\theta}_2=\left(k_2,\underset{\sim}{\pi}_2,\underset{\sim}{\mu}_2,\underset{\sim}{\sigma}_2^2\right)$, is the same as $\underset{\sim}{\theta}_1$.

## 4. Estimation some performance measures via the BD-MCMC output

Given a sample realization of the MCMC output and a sample from $f\left(\lambda\left|\underset{\sim}{t}\right.\right)$ and $f\left(p\left|\underset{\sim}{x}\right.\right)$ of equal size, we can estimate performance measures. For instance, given sample data, we would like to assess whether or not the model is stable. The system is stable if and only if the traffic intensity, $\rho$, is less than one, see e.g. Allen (1990), where for our queuing model $\rho=\rho_1+p\rho_2$, see e.g. Salehi-Rad and Mengersen (2002). Thus, the estimation of the probability of having a stable system is

$$P\left(\rho<1\left|\underset{\sim}{t},\underset{\sim}{x},\underset{\sim}{s}_1,\underset{\sim}{s}_2\right.\right)\approx\frac{1}{N}\#\left\{\rho^{(n)}<1\right\} \tag{9}$$

where

$$\rho^{(n)} = \lambda^{(n)} \left( \sum_{i=1}^{k_1^{(n)}} \pi_{1i}^{(n)} \mu_{1i}^{\prime(n)} + p^{(n)} \sum_{i=1}^{k_2^{(n)}} \pi_{2i}^{(n)} \mu_{2i}^{\prime(n)} \right) \tag{10}$$

Note that $\mu_{1i}^{\prime(n)}$ and $\mu_{2i}^{\prime(n)}$ are the mean service and re-service time, respectively. With some algebraic computation they are given by

$$\mu_{ji}^{\prime(n)} = \mu_{ji}^{(n)} + \sigma_{ji}^{(n)} \frac{\phi\left(\mu_{ji}^{(n)} / \sigma_{ji}^{(n)}\right)}{\Phi\left(\mu_{ji}^{(n)} / \sigma_{ji}^{(n)}\right)}, \qquad j = 1,2$$

in which $\left\{\left(k_1^{(n)}, \underset{\sim}{\pi}_1^{(n)}, \underset{\sim}{\mu}_1^{(n)}, \underset{\sim}{\sigma}_1^{2(n)}\right)\right\}_{n=1}^{N}$ , $\left\{\left(k_2^{(n)}, \underset{\sim}{\pi}_2^{(n)}, \underset{\sim}{\mu}_2^{(n)}, \underset{\sim}{\sigma}_2^{2(n)}\right)\right\}_{n=1}^{N}$ are the samples of size N obtained from the BD-MCMC algorithm, also $\left\{\lambda^{(n)}\right\}_{n=1}^{N}$ and $\left\{p^{(n)}\right\}_{n=1}^{N}$ are the samples of size N generated from the posterior distribution of $\lambda$ and $p$, respectively. Furthermore, a consistent estimator of the traffic intensity is

$$E\left(\rho \middle| \underset{\sim}{t}, \underset{\sim}{x}, \underset{\sim}{s}_1, \underset{\sim}{s}_2\right) \approx \frac{1}{N} \sum_{n=1}^{N} \rho^{(n)}$$

where $\rho^{(n)}$ is given in (10). Thus, by equilibrium assumption the estimation of the traffic intensity is

$$E\left(\rho \middle| \underset{\sim}{t}, \underset{\sim}{x}, \underset{\sim}{s}_1, \underset{\sim}{s}_2, \rho < 1\right) \approx \frac{1}{M} \sum_{m:\rho^{(m)}<1} \rho^{(m)} \tag{11}$$

where $M = \#\left\{\rho^{(m)} < 1\right\}$.

Now, by considering stability and using the MCMC estimations of the system parameters, i.e. $\underset{\sim}{\theta} = \left(\lambda, \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, p\right)$, $\rho_1, \rho_2, \delta_1$ and $\delta_2$ can be estimated. Then we can estimate the mean system size, formula (2) and (3), mean busy period, formula (4), and the probability of idle period of the system, formula (5), as you see in below example. For this, in the next section, we describe an example for estimating these performance measures.

## 5. Applied example

In this section, we analyze the data which has been obtained from a hospital. Patients are skin patients who arrive with a Poisson process and stay in the MQ. At first, a doctor treats these patients in the MQ. Some of them after treating, with probability $p$, need to have a corrective surgery and should go to a waiting room (surgery room). Then, after treating all patients in the MQ; the doctor goes to the waiting room for surgery these patients. Interarrival, service (treat) and re-service (corrective surgery) times of 165 patients in which 20 of them need re-service (corrective surgery) have been recorded from 9:00 AM to 16:00 PM in the period of six days.

The model for controlling the patients queue is the same model we introduced in section 2. In this queuing system, we have used the proposed BD-MCMC method for the service and re-service data.

Figure 2, illustrates the changes in the value of $k_1$ and $k_2$ against iterations. In fact, the posterior probabilities for different values of $k_1$ are (approximately) $P\left(k_1=1\middle|\underset{\sim}{S}_1\right)\approx 0.191$, $P\left(k_1=2\middle|\underset{\sim}{S}_1\right)\approx 0.341$, $P\left(k_1=3\middle|\underset{\sim}{S}_1\right)\approx 0.203$, $P\left(k_1=4\middle|\underset{\sim}{S}_1\right)\approx 0.182$, $P\left(k_1=5\middle|\underset{\sim}{S}_1\right)\approx 0.06$, with other values of $k_1$ having posterior probabilities less than 0.03. Thus, although there is some uncertainty here, having two truncated Normal mixture components appears to have been spotted by the method. Also, the posterior probabilities for different values of $k_2$ are (approximately) $P\left(k_2=1\middle|\underset{\sim}{S}_2\right)\approx 0.421$, $P\left(k_2=2\middle|\underset{\sim}{S}_2\right)\approx 0.356$, $P\left(k_2=3\middle|\underset{\sim}{S}_2\right)\approx 0.145$ with other values of $k_2$ having posterior probabilities less than 0.04. Thus, for re-service distribution, having the single trancated Normal distribution appears to have been spotted by the method.

Figure 3, in the left, shows the histogram of 165 service times and in the right shows the histogram of 43 re-service times which have obtained from the data and the predictive density of them obtained after a run of 10000 iterations. For estimating the general density functions of service and re-service times, we have used the mixture of truncated Normal distributions on interval $(-\infty,0)$ with the BD-MCMC algorithm by applying the formula (8).

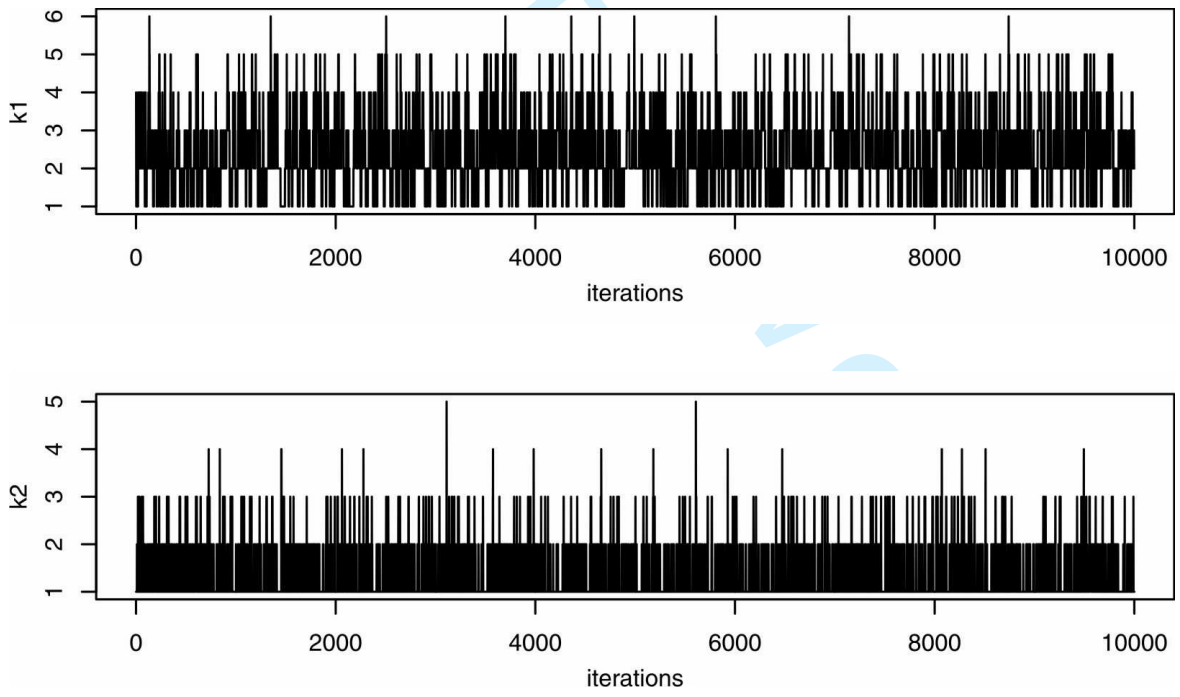


Fig. 2. Changes in $k_1$ against the 10000 iterations for the service time data (top) and the changes in $k_2$ for re-service time data (bottom).

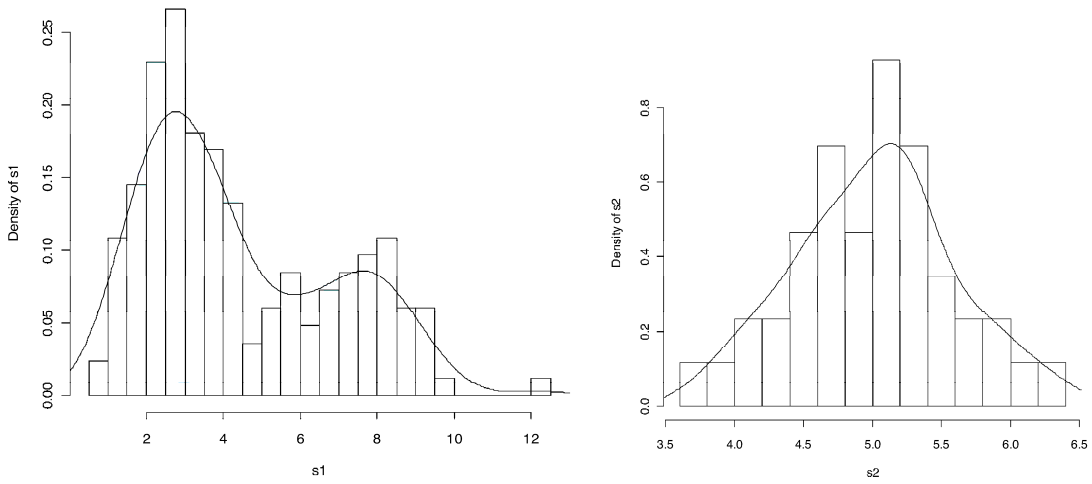


Fig. 3. Histogram of service time data and estimated service time density (left) and histogram of re-service time data and estimated re-service time density (right).

Given the observed data $(\underset{\sim}{t},\underset{\sim}{x},\underset{\sim}{s}_1,\underset{\sim}{s}_2)$, we now estimate the probability of having a stable system, $E(X_n)$, $E(Y_n)$, $\mathrm{E}[\text{busy period}]$ and $\pi^*$ by using these estimations, see (9), (11), (2), (3), (4) and (5) respectively.

Table 1

Estimations of probabilities that system is stable and, mean number of customers in the system, mean busy period and probability of idle period.

| $P(\rho<1\|data)$ | $E(\rho\|data)$ | $E(X_n)$ | $E(Y_n)$ | $\mathrm{E}[\text{busy period}]$ | $\pi^*$ |
|---|---|---|---|---|---|
| 0.9541 | 0.3708 | 43.7853 | 46.2315 | 6.4678 | 0.1346 |

According to above table, the estimations of $P(\rho<1|data)$, $E(\rho|data)$, $E(X_n)$, $E(Y_n)$, $\mathrm{E}[\text{busy period}]$ and $\pi^*$ have obtained for $k_1=2$ and $k_2=1$. Of course, we should say that for $k_j>2(j=1,2)$, the computations of $E(X_n)$ and $E(Y_n)$ are so complicated.

## 6. Further work and conclusion

A Bayesian approach to statistical inference and prediction in the M/G/1 queuing system with optional second service under a special policy has been presented. We have developed a density estimation method based on mixtures of truncated Normal distributions as a general semi-parametric model for a non-negative continuous distribution in order to approximate the general distribution of service and re-service times. To make inference on the service and re-service parameters, we have implemented an MCMC algorithm based on births and deaths of mixture components making use of the BD-MCMC technique proposed by Stephens (2000a).

Some important measures of system, such as the mean system size, mean busy period and probability of idle period, has been predicted. We have illustrated this methodology with real data.

We could consider this queuing model with a Bayesian approach under another discipline, see Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004), they considered this queuing model from view point classical queuing theory under three disciplines. Furthermore, as a referee pointed out, an interesting extension for our model could be dependence between the service and re-service times. Also, it could be considered the probability of re-service depends of specific factors. It is ongoing.

Although, in this article, we have used a mixture of truncated Normal distributions to model service and re-service times, also, it could considered more standard mixtures based on Skewed truncated Normal distribution or Generalized Exponential distribution. Density estimation problems for skewed and multimodal contexts seem suitable for our approach. We could also consider classes of phase-type distributions; see Ausin et al. (2004). There is another Bayesian approaches that we have not considered here, see Rios et al. (1998), Ausin et al. (2007b), they used a mixture of Erlang distributions for Bayesian approaches. For our queuing model, we can also use a mixture of Erlang distributions. But using a mixture of truncated Normal distributions has some advantages. For example, in our Bayesian approaches, almost, all posterior distributions can be obtained. Also, for some special cases maybe a mixture of truncated Normal distributions will be more suitable (fitter) than other models.

## **APPENDIX:** PROOF OF THEOREM 2.1

To find $E(X_n)$ and $E(X_n)$, we can apply two methods. First method is using a direct method to obtain these means and another one is using the probability generating function. We apply the second. For this, we use the joint p.g.f. of $(X_n, Y_n)$ that is

$$P(u,v) = \frac{(1-p+pv)B_1^*\left[\lambda(1-u)\right]\left[R(v) - G^*(u,p,\lambda) + (1-u)G^*(0,p,\lambda)\right]}{(1-p+pv)B_1^*\left[\lambda(1-u)\right] - u}\pi_{0\bullet} \quad (12)$$

in which

$$G^*(u,p,\lambda) = \Psi\left(1-p+pB_2^*\left[\lambda(1-u)\right]\right), \quad (13)$$

$$R(v) = \Psi(1-p+pv)$$

$$\Psi(u) = uB_1^*\left[\lambda(1-\Psi(u))\right], \quad (14)$$

and $B_1^*(.)$ and $B_2^*(.)$ are the LSTs of the distribution functions of the service and re-service times, respectively. $\pi_{0\bullet}$ is the probability that MQ is empty and is equal to

$$\pi_{0\bullet} = \Pr(X_n = 0) = (1-\rho_1)^2\left[p\rho_2 + (1-\rho_1)G^*(0,p,\lambda)\right]^{-1}. \quad (15)$$

For more details, see Salehi-Rad and Mengersen (2002, page 1718).

Based on formula (12) and according to $R(1)=\psi(1)=1$ and some simple computations, the marginal p.g.f.'s of $X_n$ and $Y_n$ in steady state are

$$P(u)=\frac{B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]}{B_1^*\left[\lambda(1-u)\right]-u}\pi_0. \tag{16}$$

$$P(v)=\frac{(1-p+pv)\left[1-R(v)\right]}{p(1-v)}\pi_0.$$

respectively.

Now, we only obtain $E(X_n)$ and for $E(Y_n)$ is the same. For finding $E(X_n)$ by the marginal p.g.f. of $X_n$, first, we derivative (16) with respect to u and set $u=1$. We obtain the derivation of (16) by means of a sequence of three following lemmas.

**Lemma 1.** If $h(x)=f(x)/g(x)$ and there are the first and second derivations of nominator and dominator of $h(x)$ in $x=a$ and $f(x)_{|x=a}=g(x)_{|x=a}=0$, we have

$$h'(x)_{|x=a}=\frac{f''(x)_{|x=a}\,g'(x)_{|x=a}-g''(x)_{|x=a}\,f'(x)_{|x=a}}{2\left(g'(x)_{|x=a}\right)^2}$$

**Lemma 2.** According to the (13) and (14) formulas, we have

$$i)\quad \frac{d}{du}\left(G^*(u,p,\lambda)\right)_{|u=1}=\frac{p\rho_2}{1-\rho_1} \tag{17}$$

$$ii)\quad \frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right)_{|u=1}=\frac{p}{1-\rho_1}\left[\lambda^2\delta_2+\rho_2^2+\frac{p\rho_2^2}{1-\rho_1}\left(2\rho_1+\frac{\lambda^2\delta_1+\rho_1^2}{1-\rho_1}\right)\right] \tag{18}$$

**Proof.** By using (13) and (14), we have

$$G^*(u,p,\lambda)=\left(1-p+pB_2^*\left[\lambda(1-u)\right]\right)B_1^*\left[\lambda\left(1-G^*(u,p,\lambda)\right)\right]$$

now, the derivative of $G^*(u,p,\lambda)$ with respect to u in given by

$$\frac{d}{du}\left(G^*(u,p,\lambda)\right)=p\frac{d}{du}\left(B_2^*\left[\lambda(1-u)\right]\right)B_1^*\left[\lambda\left(1-G^*(u,p,\lambda)\right)\right]$$
$$+\left(1-p+pB_2^*\left[\lambda(1-u)\right]\right)\frac{d}{du}\left(G^*(u,p,\lambda)\right)\lambda\int_0^\infty se^{-\lambda\left(1-G^*(u,p,\lambda)\right)s}dB_1(s)$$

by setting $u=1$ and some simple computations, we have

$$\frac{d}{du}\left(G^*(u,p,\lambda)\right)=p\rho_2+\rho_1\frac{d}{du}\left(G^*(u,p,\lambda)\right)$$

by solving this equation, the relation (i) is yield. Also, for proving (ii), we obtain

$$\frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right) = p\frac{d^2}{du^2}\left(B_2^*\left[\lambda(1-u)\right]\right)B_1^*\left[\lambda\left(1-G^*(u,p,\lambda)\right)\right]$$
$$+2p\frac{d}{du}\left(B_2^*\left[\lambda(1-u)\right]\right)\frac{d}{du}\left(G^*(u,p,\lambda)\right)\lambda\int_0^\infty se^{-\lambda\left(1-G^*(u,p,\lambda)\right)s}dB_1(s)$$
$$+\left(1-p+pB_2^*\left[\lambda(1-u)\right]\right)\left[\frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right)\lambda\int_0^\infty se^{-\lambda\left(1-G^*(u,p,\lambda)\right)s}dB_1(s)\right.$$
$$\left.+\left(\frac{d}{du}\left(G^*(u,p,\lambda)\right)\right)^2\lambda^2\int_0^\infty s^2e^{-\lambda\left(1-G^*(u,p,\lambda)\right)s}dB_1(s)\right]$$

now, by setting $u=1$, we have

$$\frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right) = p\left(\lambda^2\delta_2+\rho_2^2\right)+2p\rho_2\left(\frac{p\rho_2}{1-\rho_1}\right)\rho_1$$
$$+\rho_1\frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right)+\left(\frac{p\rho_2}{1-\rho_1}\right)^2\left(\lambda^2\delta_1+\rho_1^2\right)$$

by solving this equation, we can obtain (ii).

**Lemma 3.** Under our queuing model conditions, we have

$$i)\ \frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]\right)_{|u=1} = -\frac{(1-\rho_1)}{\pi_{0\bullet}} \qquad (19)$$

$$ii)\ \frac{d^2}{du^2}\left(B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]\right)_{|u=1}$$
$$= -\frac{2\rho_1(1-\rho_1)}{\pi_{0\bullet}}-\frac{p}{1-\rho_1}\left[\lambda^2\delta_2+\rho_2^2+p\rho_2^2\left(\frac{\lambda^2\delta_1+\rho_1(2-\rho_1)}{1-\rho_1}\right)\right] \qquad (20)$$

$$iii)\ \frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]-u\right)_{|u=1} = -(1-\rho_1) \qquad (21)$$

$$iv)\ \frac{d^2}{du^2}\left(B_1^*\left[\lambda(1-u)\right]-u\right)_{|u=1} = \lambda^2\delta_1+\rho_1^2 \qquad (22)$$

**Proof.** For (i) we have

$$\frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]\right)$$
$$=\frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]\right)\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]$$
$$+B_1^*\left[\lambda(1-u)\right]\left[-\frac{d}{du}\left(G^*(u,p,\lambda)\right)-G^*(0,p,\lambda)\right]$$

by setting $u=1$ and using (17), we obtain

$$\frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]\right)_{|u=1} = -\left[\frac{p\rho_2}{1-\rho_1}+G^*(0,p,\lambda)\right]$$

now, with some computations and using (15), we can obtain (i).

Also, for proving (ii), from (i), we have

$$\frac{d^2}{du^2}\left(B_1^*\left[\lambda(1-u)\right]\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]\right)$$
$$=\frac{d^2}{du^2}\left(B_1^*\left[\lambda(1-u)\right]\right)\left[1-G^*(u,p,\lambda)+(1-u)G^*(0,p,\lambda)\right]$$
$$+2\frac{d}{du}\left(B_1^*\left[\lambda(1-u)\right]\right)\left[-\frac{d}{du}\left(G^*(u,p,\lambda)\right)-G^*(0,p,\lambda)\right]$$
$$+B_1^*\left[\lambda(1-u)\right]\left[-\frac{d^2}{du^2}\left(G^*(u,p,\lambda)\right)\right]$$

by setting $u=1$, and using equations in Lemma 2, our proof is completed. The proofs of (iii) and (iv) are similar.

Now, for obtaining $E(X_n)$ by marginal p.g.f. of $X_n$, note that, the numerator and denominator of the marginal p.g.f., for $u=1$, are zero. So, we can use Lemma 1 to obtain $E(X_n)$. For this, by using Lemma 2 and Lemma 3 we can easily obtain $E(X_n)$ and for $E(Y_n)$ is the same.

**Acknowledgements**

We wish to thank Dr. N. Nematolahi for helpful discussions and comments on this work, and Prof. G. Gholami for discussing Bayesian approaches. We would like to thanks three anonymous referees for valuable comments and suggestions.

**References**

[1] Allen, A.O., (1990), Probability, Statistics and Queueing Theory with Computer Science Applications. Academic Press, Boston.

[2] Armero, C., Bayarri, M.J., (1994a), Bayesian prediction in M/M/1 queues. Queueing Systems 15, 401–417.

[3] Armero, C., Bayarri M.J., (1994b), Prior assessments for prediction in queues. Statistician 43(1), 139–153.

[4] Ausin, M.C., Wiper, M.P., (2007a), Bayesian control of the number of servers in a GI/M/c queueing system, J. Statist. Planning and Inference, 137, 3043-3057.

[5] Ausin, M.C., Wiper, M.P., Lillo, R.E., (2007b), Bayesian prediction of the transient behaviour and busy period in short- and long-tailed GI/G/1 queueing systems, Computational Statistics & Data Analysis, 52, 1615-1635.

[6] Ausin, M.C.,Wiper, M.P., Lillo, R.E., (2004), Bayesian estimation for the M/G/1 queue using a phase type approximation. Journal of Statistical Planning and Inference, 118, 83–101.

[7] Cappé, O., Robert, C.P., Rydén, T., (2003), Reversible jump MCMC converging to birth-and-death MCMC and more general continuous time samplers. J. Roy. Statist. Soc. Ser. B 65, 679–700.

[8] Diebolt, J., Robert, C.P., (1994), Estimation of finite mixture distributions through Bayesian sampling. J. Roy. Statist. Soc. B 56, 363–375.

[9] Green, P., (1995), Reversible jump MCMC computation and Bayesian model determination. Biometrika, 82, 711–732.

[10] Richardson, S., Green, P., (1997), On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statististical Society, B, 59, 731-792.

[11] Rios, D., Wiper, M.P., Ruggeri, F., (1998), Bayesian analysis of M/Er/1 and M/Hk/1 queues. Queueing Systems 30, 289–308.

[12] Robert, C.P., (1996), Mixtures of distributions: inference and estimation. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 441–464.
[13] Salehi-Rad, M.R., Mengersen, K., (2002), Reservicing some customers in M/G/1 queues, under two disciplines, Advances in Statestics, Combinatorics and Related Areas, World Scientific Publishing, New Jersey, pp. 267-274.
[14] Salehi-Rad, M.R., Mengersen, K., Shahkar, G. H., (2004), Reservicing some customers in M/G/1 queues, under three disciplines, International Journal of mathematics and mathematical sciences, no. 32, 1715-1723.
[15] Stephens, M., (2000a), Bayesian analysis of mixture models with an unknown number of components an alternative to reversible jump methods. Ann. Statistics, 28, 40-74.
[16] Stephens, M., (2000b), Dealing with label switching in mixture models: J. Royal Statistical Sociaty, B, v. 62, no. 4, 795–809.
[17] Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions. The Annals of Statistics, 22, 1701-1728.
[18] Wiper, M.P., Rios, D., Ruggeri, F., (2001), Mixtures of gamma distributions with applications. J. Computer Graphical Statistics, 10, 440–454.

May 11, 2011

Professor N. Balakrishnan

Editor of Chief,

Communications in Statistics-Simulation and Computation

Dear Professor Balakrishnan

This is with reference to your May 9, 2011 Email about the referee`s comment on the manuscript ID: LSSP-2011-0087.R1 entitled "Bayesian inference and prediction in an M/G/1 with optional second service" co-authored by Abdolreza Mohammadi and Mohammad Reza Salehi-Rad for Communications in Statistics-Simulation and Computation. We respond below to the reviewer(s) comments. Thanks in advance for your time.

Looking forward to hearing from you.

Best Regards,

Mohammad Reza Salehi-Rad
Assistant Prof. of Statistics
Department of Statistics
Allameh Tabataba'i University
Tehran, Iran.

Abdolreza Mohammadi
PhD student
Johann Bernoulli Institute
University of Groningen
Groningen, Netherlands.

**Response to Referees:**

Reviewer: 1
Comments to the reviewer

Although in the steady state, the distribution of the variables Xn (for n=1,2,…) are the same, the variables themselves are different, and can take different values. Therefore it is more precise to call them with different names.