Bayesian Analysis of the Stationary MAP$_2$

P. Ramírez-Cobo[*], R. E. Lillo[†], and M. P. Wiper[‡]

Abstract. In this article we describe a method for carrying out Bayesian estimation for the two-state stationary Markov arrival process (MAP_2), which has been proposed as a versatile model in a number of contexts. The approach is illustrated on both simulated and real data sets, where the performance of the MAP_2 is compared against that of the well-known $MMPP_2$. As an extension of the method, we estimate the queue length and virtual waiting time distributions of a stationary $MAP_2/G/1$ queueing system, a matrix generalization of the $M/G/1$ queue that allows for dependent inter-arrival times. Our procedure is illustrated with applications in Internet traffic analysis.

MSC 2010 subject classifications: 62F15, 62M05, 60K20.

Keywords: phase-type distributions, Markov modulated Poisson process ($MMPP$), Identifiability, canonical representation, Gibbs sampler, steady-state distributions.

1 Introduction

The *Markov arrival process* (MAP) was introduced by Neuts (1979) as a versatile point process where the times between occurrences of events (from now on, inter-event times) are phase-type (PH) distributed (see Neuts (1974)) and correlated. Applications of the MAP have been proposed in diverse areas such as teletraffic modeling, reliability, finance and climatology. For examples, see Okamura et al. (2009), Casale et al. (2010), Cheung and Landriault (2010), Cheung and Runhuan (2013), Montoro-Cazorla and Pérez-Ocón (2014), Ramírez-Cobo et al. (2014b) and Rodríguez et al. (2015).

However, the MAP has mainly been studied from a queueing theory perspective. A number of variants of the $MAP/G/1$ queueing system have been considered in e.g. Wu et al. (2011), Zhang and Hou (2011), Xue and Alfa (2011), Chaudhry et al. (2013), Dudina et al. (2013), and Banerjee et al. (2015). Also, Ramaswami (1980) derived a detailed theoretical analysis of the single-server queue where the arrival process is governed by a MAP with batch arrivals, which was studied later by Lucantoni (1991, 1993), where different numerical algorithms for computing the steady-state solutions were given.

While performance analysis for models incorporating MAPs is a well developed area, less progress has been made on statistical estimation for such processes. Probably the major difficulty for inference for MAPs is that they suffer from *identifiability* problems

[*]Department of Statistics and Operations Research, University of Cádiz, Spain, pepa.ramirez@uca.es

[†]Department of Statistics, University Carlos III of Madrid, Spain and Institute of Financial Big Data UC3M, Spain, lillo@est-econ.uc3m.es

[‡]Department of Statistics, University Carlos III of Madrid, Spain, mwiper@est-econ.uc3m.es

so that many different *MAP* parameterizations produce the same joint density for any observed sequence of inter-event times when the underlying states of the process are not observed (Ramírez-Cobo et al., 2010b). In the context of statistical inference, this implies that it is not sensible to estimate the individual parameters of the *MAP* given a sample of inter-event time data, since there are infinitely many parameterizations which give the same density for the observed process. Some papers have investigated approaches to inference for *MAP*s, either from a moment matching method perspective or via maximum likelihood, see e.g. Riska et al. (2002) or Kriege and Buchholz (2010) and the references given therein for a recent review. However, in most of these articles, with the exceptions of Telek and Horváth (2007) and Bodrog et al. (2008) the issue of identifiability is not accounted for.

Although there has been some work on Bayesian estimation for a specific, identifiable subclass of *MAP*s, that is the Markov modulated Poisson process (*MMPP*), see Scott (1999), Scott and Smyth (2003), Fearnhead and Sherlock (2006), as far as we know, Bayesian inference for the general *MAP* has not previously been considered. Recently however, an identifiable representation for the simplest, two-state stationary *MAP* (MAP_2), has been found by Bodrog et al. (2008). Thus, using this canonical representation of the MAP_2 the identifiability problems are removed. Therefore, the first objective of this paper is to undertake Bayesian estimation of the stationary MAP_2.

Concerning the $MAP/G/1$ queueing system, most of works are focused on theoretical aspects and not on inferential issues, with the exception, to the best of our knowledge, of Riska et al. (2002). However, in teletraffic contexts where the *MAP* is used as an arrival model, there will typically be parameter uncertainty present and it will be of interest to estimate probabilities of congestion, queue lengths, waiting time distributions etc. based on sample data. Similarly, the study of ruin problems in insurance is directly related to the analysis of queueing systems, see Prabhu (1998), Asmussen and Albrecher (2010) or Ramírez-Cobo et al. (2010a). Therefore, given real insurance claim data (which are commonly dependent), it will be important to make inference for the $MAP/G/1$ queue in order to obtain an estimation of the ruin probability. Thus, the second objective of this work is to obtain numerical predictions of the steady state distributions of the $MAP_2/G/1$ queue which, as far as we know, is an unexplored problem in the literature. Our MCMC (Markov chain Monte Carlo) algorithm for Bayesian inference for the MAP_2 is combined with results from queueing theory to obtain numerical predictions of the waiting time and queue length distributions of the system.

The rest of this paper is organized as follows. In Section 2 we review the definition and key properties of the stationary MAP_2. Special emphasis is put on comparing the MAP_2 with the two-state *MMPP* ($MMPP_2$), the most considered subclass of *MAP* in the literature. In Section 3 we introduce a MCMC algorithm for Bayesian inference for the stationary MAP_2. Our approach is illustrated in detail with simulated and real data sets. In Section 4 we examine the $MAP_2/G/1$ queueing system and show how our approach for estimating the MAP_2 can be used to approximate the steady state distributions of interest. Our results are then applied to a real example of Internet traffic arrivals. Conclusions and possible extensions to this work are considered in Section 5.

2 The stationary MAP$_2$

This section summarizes the main properties of the stationary MAP_2, with special emphasis on the identifiability problem and differences between the MAP_2 and $MMPP_2$ processes.

2.1 The MAP and the MAP$_2$

Formally, a (m state) MAP is a doubly stochastic process $\{J(t), N(t)\}$, where $J(t)$ represents an irreducible, continuous-time, Markov process with finite state space $\mathcal{S} = \{1, 2, \ldots, m\}$, and $N(t)$ is a counting process that records the number of events occurring in $(0, t]$.

The MAP evolves as follows. An initial state $i_0 \in \mathcal{S}$ is generated according to some initial probability vector $\boldsymbol{\alpha}$ and at the end of an exponentially distributed sojourn time in state $i \in \mathcal{S}$, with mean $1/\lambda_i$, one of two types of state transition can occur. Firstly, with probability $0 \le p_{ij1} \le 1$, a single event occurs and the MAP enters a state $j \in \mathcal{S}$, which may be the same as the previous state. Otherwise, with probability $0 \le p_{ij0} \le 1$, no event occurs and the MAP enters a different state $j \ne i$. Clearly, we have that for $i \in \mathcal{S}$, then

$$\sum_{j=1, j\ne i}^{m} p_{ij0} + \sum_{j=1}^{m} p_{ij1} = 1.$$

A MAP can thus be expressed in terms of the initial probability vector $\boldsymbol{\alpha}$ and the parameters $\{\boldsymbol{\lambda}, P_0, P_1\}$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, and P_0 and P_1 are $m \times m$ transition probability matrices with ij'th elements p_{ij0} and p_{ij1}, respectively.

From now on, we shall concentrate on the specific properties of the MAP_2, that is the two state MAP with $m = 2$ and state space $\mathcal{S} = \{1, 2\}$. Good reviews of the theoretical properties of the general MAP can be seen in e.g. Lucantoni (1993), Latouche and Ramaswami (1999), and Chakravarthy (2001).

As with any MAP, instead of transition probability matrices, the MAP_2 can be characterized in terms of rate (or intensity) matrices. For the MAP_2, these matrices are $\{D_0, D_1\}$, where

$$D_0 = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{120} \\ \lambda_2 p_{210} & -\lambda_2 \end{pmatrix}, \quad D_1 = \begin{pmatrix} \lambda_1 p_{111} & \lambda_1(1 - p_{120} - p_{111}) \\ \lambda_2 p_{211} & \lambda_2(1 - p_{210} - p_{211}) \end{pmatrix}. \tag{1}$$

Intuitively, D_1 (D_0) can be thought of as governing transition rates that do (do not) generate occurrences of events. $D = D_0 + D_1$ is the infinitesimal generator of the underlying Markov process $J(t)$ with stationary probability vector $\boldsymbol{\pi} = (\pi, 1 - \pi)$ and computed as $\boldsymbol{\pi} D = \mathbf{0}$. The stationary version of the process which we shall consider in this work is obtained when $\boldsymbol{\alpha} = \boldsymbol{\pi}$.

The stationary MAP_2 can be viewed as a Markov renewal process. Indeed, let S_r denote the state of the MAP_2 at the time of the rth event, and let T_r denote the time between the $(r-1)$th and rth events. Then $\{S_{r-1}, \sum_{i=1}^{r} T_i\}_{r=1}^{\infty}$ is a Markov renewal

process, and in particular, $\{S_r\}_{r=0}^{\infty}$ is a Markov chain with transition matrix $P^{\star}$ given by

$$P^{\star} = (I - P_0)^{-1} P_1. \tag{2}$$

From the practical viewpoint, it will often be the case that only a sequence of inter-event times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ are observed and that the corresponding state sequence of the MAP_2 is not. Therefore, it is important to consider the distribution of the variable T, representing the time between two successive events (inter-event time) in the stationary version of a MAP_2. It is known that T is phase-type distributed with representation $\{\boldsymbol{\phi}, D_0\}$, see e.g. Telek and Horváth (2007), and therefore the moments of T can be computed as

$$\mu_n = E(T^n) = n! \boldsymbol{\phi} \left(-D_0\right)^{-n} \mathbf{e}^T, \tag{3}$$

where $\boldsymbol{\phi} = (\phi, 1 - \phi)$ is the probability distribution satisfying

$$\boldsymbol{\phi} P^{\star} = \boldsymbol{\phi}, \tag{4}$$

and $\mathbf{e}$ is a row vector with all its components equal to one. The inter-event times in a stationary MAP_2 are known to be correlated and in e.g. Bodrog et al. (2008), it is shown that the autocorrelation coefficient of lag h, say R_h, is given by

$$R_h = \gamma^h \; \frac{\mu_2 - 2\mu_1^2}{2(\mu_2 - \mu_1^2)}, \quad \text{for } k > 0, \tag{5}$$

where $-1 \leq \gamma < 1$ is one of the two eigenvalues of the transition matrix $P^{\star}$ (as $P^{\star}$ is stochastic, then necessarily the other eigenvalue is equal to 1). Note that the specific structure in (5) implies geometric decay of the autocorrelation function, a property that is also shared by the general m-state MAPs, see Hervé and Ledoux (2013).

Finally, the likelihood function for a sequence of observed inter-event times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ is given by

$$f(\mathbf{t}|D_0, D_1) = \boldsymbol{\phi} e^{D_0 t_1} D_1 e^{D_0 t_2} D_1 \ldots e^{D_0 t_n} D_1 \mathbf{e}^T, \tag{6}$$

see for example, Chakravarthy (2001).

2.2   A canonical representation for the stationary MAP$_2$

Representation (1) of the stationary MAP_2 is known to be over-parameterized, in the same sense as in Rydén (1996) or Ramírez-Cobo et al. (2010b): a MAP_2 defined by $\{D_0, D_1\}$ is over-parametrized (or non-identifiable) if there exists a differently parametrized process $\{\widetilde{D}_0, \widetilde{D}_1\}$ such that

$$f(\mathbf{t}|D_0, D_1) = f(\mathbf{t}|\widetilde{D}_0, \widetilde{D}_1),$$

for any given sequence of inter-event times $\mathbf{t} = (t_1, \ldots, t_n)$ and for all $n \geq 1$, where $f(\mathbf{t}|D_0, D_1)$ is the observed likelihood function given by Eq. (6).

Bodrog et al. (2008) proves that any MAP_2 is completely characterized in terms of four descriptors, the first three moments and first-lag autocorrelation coefficient, $(\mu_1, \mu_2, \mu_3, R_1)$, see Eq. (3) and (5). This definition of the MAP_2 in terms of four quantities allowed them to find a unique, canonical representation for the stationary MAP_2. This canonical representation depends on the sign of the correlation parameter γ (in 5). Specifically, if $\gamma \geq 0$, then the canonical form of the MAP_2 is given by

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2), \quad P_0 = \begin{pmatrix} 0 & 1-a \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} a & 0 \\ 1-b & b \end{pmatrix}, \tag{7}$$

where, $0 < \lambda_1 \leq \lambda_2$ and $a, b \in [0, 1]$. We shall refer to this case as model $\mathcal{M}_1$. Otherwise, for those MAP_2s such that $\gamma < 0$, then their canonical form is

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2), \quad P_0 = \begin{pmatrix} 0 & 1-a \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & a \\ b & 1-b \end{pmatrix}, \tag{8}$$

where, $0 < \lambda_1 \leq \lambda_2$ and $a, b \in [0, 1]$ as earlier. We shall refer to this case as model $\mathcal{M}_2$. Note that if $a = b = 1$, then the process is neither recurrent nor identifiable.

The transition matrices (2) for the states at an event occurrence under these two models are given by

$$P^\star = \begin{pmatrix} 1-b+ab & b-ab \\ 1-b & b \end{pmatrix},$$

under model $\mathcal{M}_1$ and

$$P^\star = \begin{pmatrix} b-ab & 1-b+ab \\ b & 1-b \end{pmatrix},$$

under model $\mathcal{M}_2$, respectively.

2.3   MMPP$_2$ versus MAP$_2$

As noted in Section 1, a number of works have dealt with inference for the two-state, Markov modulated Poisson process ($MMPP_2$), which is a subclass of MAP_2s such that, in the traditional representation of the MAP_2 (1), satisfies $p_{121} = p_{211} = 0$. In particular, this implies that in the $MMPP_2$

1. the jumps from 1 to 2, or from 2 to 1, necessarily define transitions without the occurrence of an event, and

2. events only occur at self-transitions and since the elements of the diagonal of P_0 are zero then, every self-transition produces an event.

Because of the constraint $p_{121} = p_{211} = 0$, the stationary $MMPP_2$ is represented by four parameters $(\lambda_1, \lambda_2, p_{120}$ and $p_{210})$. On the contrary, the representation (1) of the general MAP_2 does not constrain any parameter to be equal to 0 (with the exception of $p_{110} = p_{220} = 0$, satisfied by all MAP_2s, including the $MMPP_2$s). This implies that in the MAP_2, events can occur either at transitions $1 \rightarrow 2$, $2 \rightarrow 1$ or self-transitions. This fact makes the MAP_2 a more complex model which, in principle, should imply more versatility. Given that the canonical representation of a general MAP_2 is also expressed

in terms of four parameters, it is of interest to study the benefits of modeling using MAP_2s instead of $MMPP_2$s. We shall address this issue below.

Firstly, the class of MAP_2s is more general than the $MMPP_2$s. Indeed, Asmussen and Koole (1993) show that stationary MAPs can approximate any stationary point process, a property which is not satisfied by $MMPP$s (as shown in Section 3 of the cited paper). The MAP_2 includes both renewal processes (phase type renewal processes as the Erlang and hyperexponential renewal process) and non-renewal processes, as is the case of the $MMPP_2$. This implies there are classes of MAP_2s that are not equivalent to any $MMPP_2$.

As an example, consider the MAP_2 defined by $\boldsymbol{\lambda} = (0.005, \ 1)$ and

$$P_0 = \begin{pmatrix} 0 & 0.02 \\ 0.0005 & 0 \end{pmatrix}, \qquad P_1 = \begin{pmatrix} 0.02 & 0.96 \\ 0.9990 & 0.0005 \end{pmatrix}. \tag{9}$$

Then, the procedure in Ramírez-Cobo et al. (2010b) can be used to show that the only potentially equivalent $MMPP_2$ to the MAP_2 given by (9) would have to be defined by $\widetilde{\boldsymbol{\lambda}} = (0.5013, \ 0.5037)$ and

$$\widetilde{P_0} = \begin{pmatrix} 0 & 1.1390 \\ 0.8606 & 0 \end{pmatrix}, \qquad \widetilde{P_1} = \begin{pmatrix} -0.1390 & 0 \\ 0 & 0.1394 \end{pmatrix},$$

which does not define a real $MMPP_2$ since $p_{120} > 1$ and $p_{111} < 0$. In order to clarify this phenomenon consider the canonical representation for (9), which can be obtained from the approach in Bodrog et al. (2008) as the model $\mathcal{M}_2$

$$\boldsymbol{\lambda} = (0.005, \ 1), \quad P_0 = \begin{pmatrix} 0 & 0.0404 \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & 0.9596 \\ 0.9994 & 6 \times 10^{-4} \end{pmatrix}.$$

It can be easily seen that in this case the autocorrelation parameter $\gamma = -0.9590$ and the first-lag autocorrelation coefficient of the inter-event times is $R_1 = -0.3219 < 0$. Not many works have been devoted to the study of the autocorrelation function of the inter-event times produced by the MAP, but for example Kang and Sung (1995) prove that for any $MMPP_2$, then $\gamma \geq 0$ and $R_h \geq 0$, for all h. On the contrary, as we show next, for certain MAP_2s the autocorrelation function R_h can take negative values, for some $h > 0$.

Consider a MAP_2 with canonical representation (7). For this case it can be seen that $\gamma = ab$ which leads to the following correlation patterns:

$$\begin{aligned} &\text{a) if } b > \frac{\lambda_1}{\lambda_2}, \text{ then } R_h > 0, \text{ for all } h > 0, \\ &\text{b) if } b < \frac{\lambda_1}{\lambda_2}, \text{ then } R_h < 0, \text{ for all } h > 0, \\ &\text{c) if } b = \frac{\lambda_1}{\lambda_2}, \text{ then } R_h = 0, \text{ for all } h > 0. \end{aligned} \tag{10}$$

If a MAP_2 has canonical representation (8), then $\gamma = -ab$ which implies that

$$\text{a) if } (1 + ab - b) < \frac{\lambda_1}{\lambda_2}, \text{ then } R_{2h-1} > 0 \text{ and } R_{2h} < 0, \text{ for all } h > 0,$$

b) if $(1 + ab - b) > \dfrac{\lambda_1}{\lambda_2}$, then $R_{2h-1} < 0$ and $R_{2h} > 0$, for all $h > 0$,

c) if $(1 + ab - b) = \dfrac{\lambda_1}{\lambda_2}$, then $R_h = 0$, for all $h > 0$. $\hspace{2cm}$ (11)

Results in Kang and Sung (1995) concerning the autocorrelation function of the inter-event times produced by a $MMPP_2$ imply that patterns described in b) of equations (10) or in any of the three cases of equations (11) will never be captured by a $MMPP_2$. And even the class of MAP_2s with a correlation pattern as in a) of equations (10) is still larger than the class of $MMPP_2$s. For example, the MAP_2 given by $\boldsymbol{\lambda} = (0.0064,\ 16.2737)$ and

$$
P_0 = \begin{pmatrix} 0 & 0.2188 \\ 0 & 0 \end{pmatrix}, \qquad P_1 = \begin{pmatrix} 0.7812 & 0 \\ 0.9992 & 0.0008 \end{pmatrix},
$$

satisfies $R_h > 0$, for all $h \geq 1$, and has no equivalent $MMPP_2$ representation. The above demonstrates the class of MAP_2s is richer than that of $MMPP_2$s and highlights the capability of the MAP_2 to capture a range of autocorrelation patterns which cannot be obtained under a $MMPP_2$.

3   Bayesian inference for the MAP$_2$

Assume now that we observe a sequence of real valued inter-event times, $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ generated from a stationary MAP_2. In this section we shall develop a Markov chain Monte Carlo algorithm for Bayesian inference in order to approximate the inter-event time distribution and to estimate the parameters of the model. The algorithm is based on the canonical forms for the MAP_2 in (7) and (8).

In order to undertake Bayesian inference, we must first introduce prior distributions for the unknown MAP_2 parameters: the exponential rates $\boldsymbol{\lambda} = (\lambda_1,\ \lambda_2)$ and transition probabilities a and b.

Unless there is specific information available a priori, it seems reasonable to apply relatively uninformative, but flexible prior distributions. Here, we adopt the approach of Gruet et al. (1999) and define the following prior structure

$$
\lambda_1 = \lambda_2 \omega, \ \text{where} \ \omega \sim \mathcal{U}(0,1),
$$

and λ_2 has a gamma prior:

$$
\lambda_2 \sim Ga\left(\alpha_2, \beta_2\right).
$$

In addition, we use independent beta (uniform) priors for a and b, $a \sim Be(\alpha_a, \beta_a)$, $b \sim Be(\alpha_b, \beta_b)$. In practice we set $\alpha_2 = 1$, $\beta_2 = 0.001$, and $\alpha_a = \alpha_b = \beta_a = \beta_b = 1$.

3.1   The complete data likelihood for the MAP$_2$

We shall base our approach on reconstruction of the complete sequence of transitions, including both the transitions that take place when an event occurs and those transitions

where there is no event but a state change occurs. Given the full history of the process up to the n'th event, then we can reconstruct the complete data likelihood function.

Let $\tilde{\mathbf{t}} = \{\tilde{t}_1, \ldots, \tilde{t}_N\}$ represent the sequence of real valued inter-transition times and $\tilde{\mathbf{s}} = \{\tilde{s}_0, \tilde{s}_1, \ldots, \tilde{s}_N\}$ the sequence of the visited states (with and without events). Note that the observed inter-event times $\mathbf{t}$ are included in the set $\tilde{\mathbf{t}}$. Define $\tilde{\mathbf{g}} = \{\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_N\}$ such that $\tilde{g}_j = 1$ (0) if the $j-$th transition occurred with (without) an event. Then the complete data likelihood is given by

$$f\left(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}, \tilde{\mathbf{g}}|\boldsymbol{\lambda}, P_0, P_1\right) = \pi_{\tilde{s}_0} \prod_{l=1}^{N} \lambda_{\tilde{s}_{l-1}} e^{-\lambda_{\tilde{s}_{l-1}}\tilde{t}_l} p_{\tilde{s}_{l-1},\tilde{s}_l,\tilde{g}_l}, \tag{12}$$

where $\boldsymbol{\pi}$ is the stationary probability vector corresponding to the underlying Markov chain, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ denotes the exponential rates, and P_0 and P_1 are the transition matrices from states 1 and 2, respectively, with elements p_{ij0} and p_{ij1}, probabilities of a transition from state i to state j with 0 or 1 events.

The complete data likelihood function in Eq. (12) can be simplified in terms of sufficient statistics. Specifically, if n_{ij} is the number of transitions with an event between states i and j, m_{12} is the number of transitions without an event, (which must all be, given the canonical parameterization, from state 1 to state 2), n_j is the number of transitions made from state j, and finally, v_j is the time spent in state j, for $i, j = 1, 2$, then from (7) and (8),

$$f(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}, \tilde{\mathbf{g}}|\omega, \lambda_2, a, b) \propto \pi_{\tilde{s}_0}(1-a)^{m_{12}} a^{n_{11}}(1-b)^{n_{21}} b^{n_{22}} \lambda_2^{n_1+n_2}\ \omega^{n_1} e^{-\lambda_2(\omega v_1 + v_2)}, \tag{13}$$

in the case of model $\mathcal{M}_1$ and

$$f(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}, \tilde{\mathbf{g}}|\omega, \lambda_2, a, b) \propto \pi_{\tilde{s}_0}(1-a)^{m_{12}} a^{n_{12}}(1-b)^{n_{22}} b^{n_{21}} \lambda_2^{n_1+n_2}\ \omega^{n_1} e^{-\lambda_2(\omega v_1 + v_2)}, \tag{14}$$

in the case of model $\mathcal{M}_2$, respectively.

3.2 Gibbs algorithm

Assuming that the MAP_2 is in one of the two canonical forms (7) or (8), we propose the following algorithm to sample the posterior distribution of the model parameters $\boldsymbol{\theta} = (\omega, \lambda_2, a, b)$.

Algorithm to estimate the MAP_2 parameters.

1. Set initial values $\boldsymbol{\theta}^{(0)} = (\omega^{(0)}, \lambda_2^{(0)}, a^{(0)}, b^{(0)})$.

2. For $k = 1, \ldots, M$ repeat:

    (a) Define $\lambda_1^{(k-1)} = \omega^{(k-1)} \lambda_2^{(k-1)}$.

    (b) Generate the states at events, $\mathbf{s}^{(k)} = (s_0^{(k)}, s_1^{(k)}, \ldots, s_n^{(k)})$, from $f(\mathbf{s}^{(k)}|\mathbf{t}, \boldsymbol{\theta}^{(k-1)})$.

(c) Complete the sequence $\tilde{\mathbf{s}}^{(k)}$ from $f(\tilde{\mathbf{s}}^{(k)}|\mathbf{t},\mathbf{s}^{(k)}\boldsymbol{\theta}^{(k-1)})$ by drawing the states when a transition without an event occurs. Obtain $\tilde{\mathbf{s}}^{(k)} = (\tilde{s}_0^{(k)},\tilde{s}_1^{(k)},\ldots,\tilde{s}_N^{(k)})$.

(d) Generate the inter-transition times, $\tilde{\mathbf{t}}^{(k)} = (\tilde{t}_1^{(k)},\ldots,\tilde{t}_N^{(k)})$, from $f(\tilde{\mathbf{t}}^{(k)}|\mathbf{t},\mathbf{s}^{(k)},\tilde{\mathbf{s}}^{(k)},\boldsymbol{\theta}^{(k-1)})$.

(e) Generate $\boldsymbol{\theta}^{(k)}$ from $f(\boldsymbol{\theta}^{(k)}|\mathbf{t},\mathbf{s}^{(k)},\tilde{\mathbf{s}}^{(k)},\tilde{\mathbf{t}}^{(k)})$.

We next describe how steps 2(b)-2(e) are undertaken at each Gibbs iteration. For simplicity of notation, we avoid the use of the superscript (k).

Step 2(b) To generate the set of states at events, $\mathbf{s} = (s_0,s_1,\ldots,s_n)$, the classic forward–backward sampler is used, see Scott (2002). In order to implement it, the likelihood of the observed inter-event times $\mathbf{t}$ given the states at events $\mathbf{s}$, is needed. Since the MAP_2 is a Markov renewal process, then the sequence of random variables representing the inter-event times, $\{T_r\}_{r=1}^{\infty}$ are conditionally independent given the sequence of states at events, $\{S_r\}_{r=1}^{\infty}$, where the distributions of $T_r \mid S_1,\ldots,S_{r-1}$ is the same as that of $T_r \mid S_{r-1}$. See Ch. 10 in Çinlar (1975) for more details. From (7) and (8), the distribution of the inter-event times conditioned on the previous states is as follows,

$$T_r|s_{r-1}=2 \quad \sim \quad Ex(\lambda_2), \tag{15}$$

$$T_r|s_{r-1}=1 \quad \sim \quad \left\{ \begin{array}{ll} Ex(\lambda_1) & \text{with probability } a, \\ Ex(\lambda_1) + Ex(\lambda_2) & \text{with probability } 1-a, \end{array} \right. \tag{16}$$

for all $r = 1,\ldots,n$, where $Ex(\lambda)$ represents an exponential variable with parameter λ. Thus, the likelihood of $\mathbf{t}$ given $\mathbf{s}$ is

$$f(\mathbf{t}|\mathbf{s},\boldsymbol{\theta}) = \phi_{s_0} \prod_{r=1}^{n} P^{\star}(s_{r-1},s_r)h\left(t_r|s_{r-1}\right), \tag{17}$$

where $P^{\star}(i,j)$ denotes the $(i,j)-$th element in the transition matrix $P^{\star}$ as in (2), and $h(\cdot|s_{r-1})$ is the density function corresponding to Eq. (15) or Eq. (16), depending on the value of s_{r-1}. From expression (17), note that we are assuming that our data start with an event. If this is not the case, the value ϕ_{s_0} should be replaced by $\boldsymbol{\pi}_{s_0}$.

Step 2(c) Consider now the generation of intermediate states to obtain the complete sequence of states, $\tilde{\mathbf{s}}$. From representations (7) and (8) there is at most one state transition between any two real events. For $r = 1,\ldots,n$, let $\tilde{S}_r = \begin{pmatrix} cc\tilde{S}_r(1,1) & \tilde{S}_r(1,2) \\ \tilde{S}_r(2,1) & \tilde{S}_r(2,2) \end{pmatrix}$ represent the number of transitions of each type that occur between real events $(r-1)$ and r, including the final transition associated with the real event. In particular, if $s_{r-1} = 2$ then there are no intermediate transitions and, we have that

$$\tilde{S}_r = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \text{ if } s_r = 2, \text{or} \quad \tilde{S}_r = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \text{ if } s_r = 1 \quad (\mathcal{M}_1,\mathcal{M}_2).$$

Otherwise, if $s_{r-1} = 1$, then either a transition without events to state 2 can occur, so that

$$\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \text{ if } s_r = 1, \text{or} \quad \tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \text{ if } s_r = 2 \quad (\mathcal{M}_1, \mathcal{M}_2),$$

or there is no intermediate transition

$$\tilde{S}_r = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \text{ if } s_r = 1 \quad (\mathcal{M}_1), \tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \text{ if } s_r = 2 \quad (\mathcal{M}_2).$$

Then, dropping the dependence on a, b for convenience, the following prior probabilities are obtained:

$$P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) = \frac{(1-a)(1-b)}{a + (1-a)(1-b)},$$

$$P\left(\tilde{S}_r = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) = 1 - \frac{(1-a)(1-b)}{a + (1-a)(1-b)}, \quad (\mathcal{M}_1)$$

$$P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 2\right) = 1,$$

or

$$P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 2\right) = \frac{(1-a)(1-b)}{a + (1-a)(1-b)},$$

$$P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 2\right) = 1 - \frac{(1-a)(1-b)}{a + (1-a)(1-b)}, \quad (\mathcal{M}_2)$$

$$P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) = 1.$$

Then, conditional on the inter-event times we have that, for model $\mathcal{M}_1$, the posterior probabilities q_1 and q_2 defined as

$$q_1 = P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1, t_r\right),$$

$$q_2 = P\left(\tilde{S}_r = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1, t_r\right)$$

are given by

$$q_1 = \frac{P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) g(t_r|\lambda_1, \lambda_2)}{P\left(\tilde{S}_r = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) g(t_r|\lambda_1, \lambda_2) + P\left(\tilde{S}_r = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \middle| s_{r-1} = 1, s_r = 1\right) f(t_r|\lambda_1)},$$

$$q_2 = 1 - q_1, \tag{18}$$

and similarly for model $\mathcal{M}_2$. In Eq. (18), $f(\cdot|\lambda)$ denotes an exponential density function with rate λ, and $g(\cdot|\lambda_1, \lambda_2)$ is the density of a sum of two exponentials with rates λ_1 and λ_2,

$$g(t|\lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} \left(1 - e^{-t(\lambda_2 - \lambda_1)}\right).$$

<u>Step 2(d)</u> To generate the complete inter-transitions times sequence, $\tilde{\mathbf{t}}$, we proceed as follows. Consider the first inter-event time, t_1. If $s_0 = 2$, then $\tilde{t}_1 = t_1$. Otherwise, if $s_0 = 1$ and no intermediate transition occurs (that is, if $s_1 = 1$), $\tilde{t}_1 = t_1$. However, if $s_0 = 1$ and an intermediate transition occurs (that is, $s_1 = 2$), then $\tilde{t}_1 = c_1$ and $\tilde{t}_2 = c_2$, such that $c_1 + c_2 = t_1$. If T_1 and C_1 represent random variables accounting for the first inter-event time, and the time at which the intermediate transition occurs, respectively, and $F_{2,2}$ denotes an F-distribution with parameters $(2,2)$ then, one has

$$\frac{T_1}{C_1} \mid s_0 = 1, \tilde{s}_0, \lambda_1, \lambda_2 \sim 1 + \frac{\lambda_2}{\lambda_1} F_{2,2}.$$

Therefore, we can generate $x \sim F_{2,2}$ and set $c_1 = \dfrac{t_1}{1 + \frac{\lambda_2}{\lambda_1} x}$ and $c_2 = t_1 - c_1$.

We can now proceed analogously to the completion of $\tilde{\mathbf{t}} = (\tilde{t}_1, \ldots, \tilde{t}_N)$.

<u>Step 2(e)</u> Finally, from the complete data likelihood functions (13)–(14) and the considered conjugate (except for ω) priors, it is easy to find the posteriors

$$
\begin{aligned}
\lambda_2|\omega, \mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Ga(\alpha_2 + n_1 + n_2,\ \beta_2 + \omega v_1 + v_2) \quad (\mathcal{M}_1, \mathcal{M}_2), \quad (19) \\
\omega|\lambda_2, \mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Ga\left(n_1 + 1,\ v_1 \lambda_2\right) \quad \text{truncated onto } [0,1] \quad (\mathcal{M}_1, \mathcal{M}_2), \\
a|\mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Be\left(\alpha_a + n_{11},\ \beta_a + m_{12}\right) \quad (\mathcal{M}_1), \\
b|\mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Be\left(\alpha_b + n_{22},\ \beta_b + n_{21}\right) \quad (\mathcal{M}_1), \\
a|\mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Be\left(\alpha_a + n_{12},\ \beta_a + m_{12}\right) \quad (\mathcal{M}_2), \\
b|\mathbf{t}, \mathbf{s}, \tilde{\mathbf{t}}, \tilde{\mathbf{s}} &\sim Be\left(\alpha_b + n_{21},\ \beta_b + n_{22}\right) \quad (\mathcal{M}_2).
\end{aligned}
$$

3.3 Model selection

The procedure to fit a MAP_2 in one of the canonical forms (7) or (8) to a given sequence of inter-event times $\mathbf{t}$, has been described in Section 3.2. However, the canonical form is unknown in practice and therefore, model selection or model averaging need to be considered. Here, we use Bayes factors for model comparison and use the approach of Chib (1995) to calculate the marginal likelihoods under each model. Consider model $\mathcal{M}_1$. Then, given point estimates of the model parameters, $\hat{\boldsymbol{\theta}} = (\hat{\omega}, \hat{\lambda}_2, \hat{a}, \hat{b})$ (for example, $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\cdot)$, the posterior mean), then $\log f(\mathbf{t}|\mathcal{M}_1)$ is approximated as follows:

$$\log f(\mathbf{t}|\mathcal{M}_1) \approx \log f(\mathbf{t}|\hat{\boldsymbol{\theta}}, \mathcal{M}_1) + \log f(\hat{\boldsymbol{\theta}}|\mathcal{M}_1) - \log f(\hat{\boldsymbol{\theta}}|\mathbf{t}, \mathcal{M}_1). \tag{20}$$

In Eq. (20), the first term denotes the likelihood function of the inter-event times, computed from Eq. (6). The second term in Eq. (20) concerns the log-prior distributions. The joint posterior distribution of the model parameters, the third term in Eq. (20), is approximated as

$$\log f(\hat{\boldsymbol{\theta}}|\mathbf{t}) \approx \frac{1}{M} \sum_{k=1}^{M} \left(\log f\left(\hat{a}|\mathbf{t}, \tilde{\mathbf{t}}^{(k)}, \mathbf{s}^{(k)}, \tilde{\mathbf{s}}^{(k)}\right) + \log f\left(\hat{b}|\mathbf{t}, \tilde{\mathbf{t}}^{(k)}, \mathbf{s}^{(k)}, \tilde{\mathbf{s}}^{(k)}\right) \right.$$

$$+ \log f\left(\hat{\omega}|\mathbf{t}, \tilde{\mathbf{t}}^{(k)}, \mathbf{s}^{(k)}, \tilde{\mathbf{s}}^{(k)}\right) + \log f\left(\hat{\lambda}_2|\mathbf{t}, \tilde{\mathbf{t}}^{(k)}, \mathbf{s}^{(k)}, \tilde{\mathbf{s}}^{(k)}\right)\right), \qquad (21)$$

where the model has been dropped for the sake of abbreviation and M is the number of runs for the Gibbs sampler in Section 3.2.

Finally, the value of $\log f(\mathbf{t}|\mathcal{M}_2)$ is computed in an analogous way which allows for the derivation of the log Bayes factor.

3.4   Numerical illustrations

In this section we illustrate the performance of the proposed MCMC algorithm for Bayesian inference of the MAP_2 on a set of simulated and real data sets.

Simulated data. In order to clarify the differences between MAP_2s and $MMPP_2$s we consider three inter-event times traces, simulated from the following models:

S1. The MAP_2 of type $\mathcal{M}_1$

$$\boldsymbol{\lambda} = (1.5244, \ 49.2821), \quad P_0 = \begin{pmatrix} 0 & 0.0657 \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0.9343 & 0 \\ 0.0194 & 0.9806 \end{pmatrix},$$

which has the equivalent $MMPP_2$ given by

$$\boldsymbol{\lambda} = (1.5264, \ 49.2801), \quad P_0 = \begin{pmatrix} 0 & 0.0669 \\ 0.0194 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0.9331 & 0 \\ 0 & 0.9806 \end{pmatrix}.$$

S2. The MAP_2 of type $\mathcal{M}_1$

$$\boldsymbol{\lambda} = (3.4364, \ 3.4453), \quad P_0 = \begin{pmatrix} 0 & 0.2454 \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0.7546 & 0 \\ 0.0021 & 0.9979 \end{pmatrix},$$

which does not have any equivalent $MMPP_2$ (from the approach in Ramírez-Cobo et al. (2010b)).

S3. The MAP_2 of type $\mathcal{M}_2$

$$\boldsymbol{\lambda} = (0.683, \ 34.6904), \quad P_0 = \begin{pmatrix} 0 & 0.0038 \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & 0.9962 \\ 0.9962 & 0.0038 \end{pmatrix}.$$

Since this MAP_2 is of type $\mathcal{M}_2$, then according to Section 2.3 it does not have any equivalent $MMPP_2$.

The generator models were randomly obtained so that they satisfied to have an equivalent $MMPP_2$ and a non-negligible autocorrelation structure for the inter-event times (S1), to lack of an equivalent $MMPP_2$ (S2), and to be a MAP_2 of type $\mathcal{M}_2$ with a non-negligible autocorrelation function (S3). The sample sizes where fixed to 1500, 500 and 1500, for S1, S2, and S3, respectively. The reason for such a choice is related to the theoretical coefficient of variation of the inter-event time distribution. Specifically,
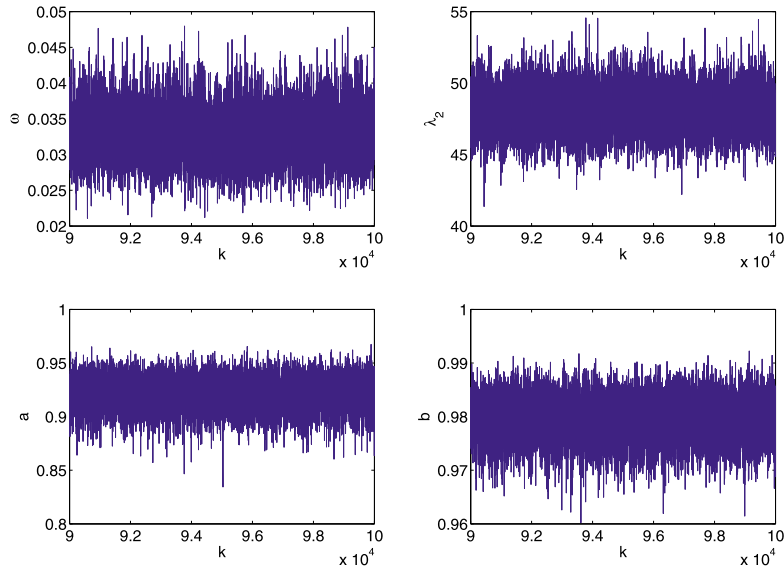
Figure 1: MCMC trace plots for simulated data set S1. The Gibbs algorithm was applied to a sample of 1500 data generated from a MAP$_2$ with parameters $\boldsymbol{\theta}$ = (0.03, 49.28, 0.9343, 0.9806).

for the first and third generator models the coefficients of variation are equal to 2.48 and 1.68, which implies that if the sample sizes are not large enough, the empirical descriptors as the mean, variance and autocorrelation coefficients might be far from the theoretical ones and thus, the estimates of the model parameters might be inaccurate. For the second model, however, the coefficient of variation is close to 1, and therefore, a smaller sample size is enough to capture the theoretical descriptors.

The Gibbs sampler described in Section 3.2 with the prior structures commented at the beginning of Section 3, was run for 100,000 iterations (with 10,000 for burn in) for each data set. A prototype code was written in Matlab and, when run on Intel Core i5 at 3.2 GHz and 16 GB of DDR3 RAM, took approximately 19 seconds to perform 1000 iterations for a sample size equal to 500 (the time increased to 56 seconds when the length of the sample was 1500). Concerning the first data set S1, Figure 1 illustrates the mixing properties of the algorithm in this case.

Tables 1 summarizes the obtained results for the simulated data set S1 under both types of MAP$_2$ ($\mathcal{M}_1$ and $\mathcal{M}_2$) and MMPP$_2$. The starting values of the MCMC, posterior mean for the model parameters, and corresponding 95% credible intervals, including those for the mean, variance, and first three autocorrelation coefficients of the inter-event times distribution are included in the table. The sample values are shown in parenthesis in the first column. Finally, the marginal likelihoods under both types of MAP$_2$ and the MMPP$_2$ are also shown in the last row of the table. Figure 2 shows the fits to the empirical CDF (cumulative distribution function) of the inter-event times

Data set **S1:**			
$\boldsymbol{\theta} = (0.03, 49.28, 0.93, 0.98)$	$MAP_2(\mathcal{M}_1)$	$MAP_2(\mathcal{M}_2)$	$MMPP_2$
$\boldsymbol{\theta}^{(0)}$	$(0.82, 7.47, 0.81, 0.44)$	$(0.16, 13.27, 0.064, 0.847)$	$(0.894, 10.294, 0.492, 0.148)$
$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{t})$	$(0.036, 48.09, 0.921, 0.979)$	$(0.354, 8.012, 0.027, 0.111)$	$(0.034, 48.117, 0.787, 0.843)$
C_ω	$[0.027,\ 0.041]$	$[0.282,\ 0.437]$	$[0.026,\ 0.034]$
C_{λ_2}	$[45.51,\ 50.68]$	$[7.59,\ 8.41]$	$[45.76,\ 50.71]$
C_a	$[0.895,\ 0.945]$	$[0.0015,\ 0.085]$	$[0.684,\ 0.906]$
C_b	$[0.972,\ 0.986]$	$[0.095,\ 0.129]$	$[0.616,\ 0.998]$
C_{μ_1} (0.164)	$[0.118,\ 0.218]$	$[0.155,\ 0.174]$	$[0.022,\ 0.733]$
C_{σ^2} (0.173)	$[0.103,\ 0.220]$	$[0.033,\ 0.551]$	$[0.002,\ 0.557]$
C_{R_1} (0.384)	$[0.353,\ 0.391]$	$[-0.007,\ 0]$	$[0.002,\ 0.39]$
C_{R_2} (0.403)	$[0.316,\ 0.357]$	$[0.0002,\ 1.5] \times 10^{-4}$	$[8 \times 10^{-8},\ 0.336]$
C_{R_3} (0.321)	$[0.281,\ 0.329]$	$[-1.3,\ -10^{-5}] \times 10^{-7}$	$[1.6 \times 10^{-11},\ 0.290]$
$\log f(\mathbf{t}|model)$	3190.6	1555.3	2941.3

Table 1: Performance of the MCMC for the simulated data set S1, under both types of MAP_2 and $MMPP_2$ (second to fourth columns, respectively). From top to bottom: starting values of the MCMC, posterior mean for the model parameters, 95% credible intervals for the model parameters, 95% credible intervals for the mean, variance, and first three autocorrelation coefficients of the inter-event times distribution, and marginal likelihoods.
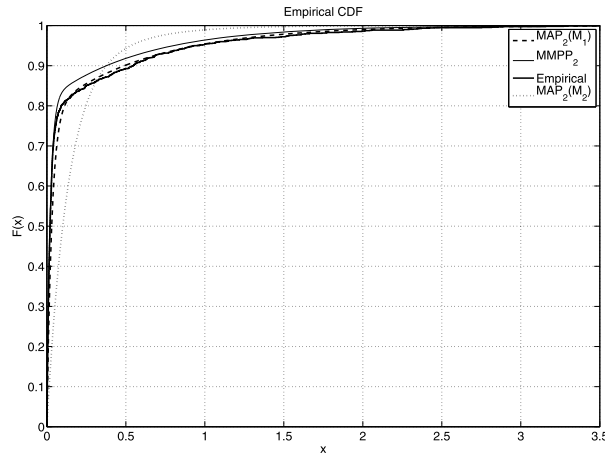
Figure 2: Empirical (thick solid line), and posterior predictive CDFs of the inter-event times via the $MAP_2(\mathcal{M}_1)$ (dashed line), the $MAP_2(\mathcal{M}_2)$ (dotted line) and the $MMPP_2$ (thin solid line) for the simulated data set S1.

provided by both types of MAP_2 and $MMPP_2$. Some comments follow from the table and figure. First, the comparison of the marginal likelihoods obtained under the three models provides evidence in favor of the generating process. Note how the $MMPP_2$ produces a marginal likelihood that is slightly lower than that of the MAP_2 of type $\mathcal{M}_1$, but much higher than the corresponding to the MAP_2 of type $\mathcal{M}_2$. From Figure 2 it is clear that the MAP_2 of type $\mathcal{M}_1$ performs better than the other approaches, the $MMPP_2$ provides a similar performance as the MAP_2 of type $\mathcal{M}_1$ and finally, the MAP_2 of type $\mathcal{M}_2$ shows the poorest behavior.

Concerning the parameters defining the generating process, it can be seen how under the generator model the credible intervals include the true values. Similarly, the intervals obtained when fitting the $MMPP_2$ capture the true parameters, except for parameter a. Regarding the sample values, all models correctly fit the mean and variance. With respect to the empirical autocorrelation function, the MAP_2 of type $\mathcal{M}_1$ outperforms the other models (note how under the MAP_2 of type $\mathcal{M}_2$ some negative values are obtained).

Consider now the analysis of the simulated data set S2, generated from a MAP_2 of type $\mathcal{M}_1$ that does not possess an equivalent $MMPP_2$. Table 2 and Figure 3 are described in analogous way as in the previous analysis for the sample S1. Some conclusions arise from the presented results. First, here again the generator model produces the highest value for the marginal likelihood, closely followed by the MAP_2 of type $\mathcal{M}_2$ and $MMPP_2$, respectively. From Figure 3 the same deduction is obtained. Note that despite having similar values for the marginal likelihoods, the estimated parameters under both types of MAP_2s do not need to be similar because representations (7) and (8) are not equal. Again, the MCMC algorithm provides reasonable estimates for the sample mean and variance (with the exception of the $MMPP_2$ for the variance). Given that the empirical

Data set **S2:**			
$\boldsymbol{\theta} = (0.997, 3.44, 0.7546, 0.9979)$	$MAP_2(\mathcal{M}_1)$	$MAP_2(\mathcal{M}_2)$	$MMPP_2$
$\boldsymbol{\theta}^{(0)}$	$(0.905, 7.234, 0.107, 0.009)$	$(0.675, 6.108, 0.441, 0.881)$	$(0.183, 7.782, 0.328, 0.874)$
$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}\|\mathbf{t})$	$(0.866, 4.315, 0.715, 0.883)$	$(0.840, 3.973, 0.683, 0.188)$	$(0.802, 8.941, 0.609, 0.673)$
C_ω	$[0.662,\ 0.998]$	$[0.542,\ 0.996]$	$[0.469,\ 0.961]$
C_{λ_2}	$[3.213,\ 5.044]$	$[3.419,\ 4.726]$	$[3.583,\ 16.446]$
C_a	$[0.505,\ 0.835]$	$[0.23,\ 0.907]$	$[0.271,\ 0.997]$
C_b	$[0.697,\ 0.998]$	$[0.014,\ 0.434]$	$[0.581,\ 0.774]$
C_{μ_1} (0.3054)	$[0.267,\ 0.311]$	$[0.267,\ 0.310]$	$[0.177,\ 0.725]$
C_{σ^2} (0.0841)	$[0.066,\ 0.097]$	$[0.069,\ 0.099]$	$[0.21,\ 0.31]$
C_{R_1} (-0.008)	$[-0.005,\ 0.006]$	$[-0.003,\ 0.003]$	$[0.061,\ 0.099]$
C_{R_2} (-0.043)	$[-0.001,\ 0.002]$	$[-3 \times 10^{-4},\ 3.5 \times 10^{-4}]$	$[2.5 \times 10^{-6},\ 0.011]$
C_{R_3} (-0.036)	$[-10^{-5},\ 7 \times 10^{-3}]$	$[-5 \times 10^{-5},\ 4 \times 10^{-5}]$	$[1.2 \times 10^{-7},\ 0.004]$
$\log f(\mathbf{t}\|model)$	114.634	111.662	104.297

Table 2: Performance of the MCMC for the simulated data set S2, under both types of MAP$_2$ and MMPP$_2$ (second to fourth columns, respectively). From top to bottom: starting values of the MCMC, posterior mean for the model parameters, 95% credible intervals for the model parameters, 95% credible intervals for the mean, variance, and first three autocorrelation coefficients of the inter-event times distribution, and marginal likelihoods.
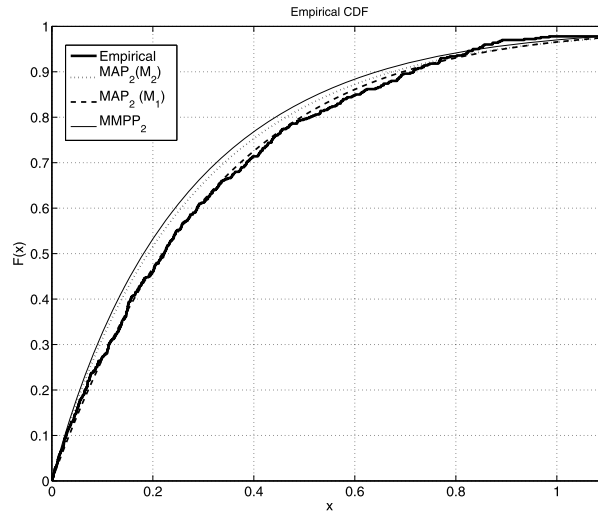
Figure 3: Empirical (thick solid line), and posterior predictive CDFs of the inter-event times via the $\mathrm{MAP}_2(\mathcal{M}_1)$ (dashed line), the $\mathrm{MAP}_2(\mathcal{M}_2)$ (dotted line) and the MMPP_2 (thin solid line) for the simulated data set S2.

autocorrelation function of the inter-event times is negligible, the estimated models are close to PH-renewal processes (subclass of MAP_2 where $R_h = 0$, for all h).

Finally, consider in Table 3 and Figure 4 the results obtained for the sample S3, which was generated by a MAP_2 of type $\mathcal{M}_2$.

Since any $MMPP_2$ is a MAP_2 of type $\mathcal{M}_1$, then as occurred in the previous simulated example, the generator model here does not have any equivalent $MMPP_2$. For this data set, again the generator model is that which presents the highest marginal likelihood followed by the MAP_2 of type $\mathcal{M}_1$ and the $MMPP_2$. The analogous conclusion is derived from Figure 4, where it can be observed how the predictive CDF by the MAP_2 of type $\mathcal{M}_2$ is almost indistinguishable from the empirical CDF. Concerning the estimation of the model parameters, note how the true values are included in the credible intervals obtained under the MAP_2 of type $\mathcal{M}_2$. The inter-event time distribution moments are well captured by the three types of models and as expected, both the MAP_2 of type $\mathcal{M}_1$ as well as the $MMPP_2$ fail in fitting the alternate autocorrelation function of the inter-event times. With the exception of R_3, the estimated MAP_2 of type $\mathcal{M}_2$ does capture the correlation coefficients.

Real data. We now illustrate the performance of the MCMC algorithm when fitting the MAP_2 to a real data set. The sample, depicted by Figure 5, reports a sequence of 823 consecutive times (in sec) between real crashes in the corresponding author's personal computer (data available from the author on request).

As in the previous results, for comparison reasons both types of MAP_2s and the $MMPP_2$ are estimated. The same prior distributions, number of iterations, and burn in

Data set **S3:**			
$\boldsymbol{\theta} = (0.019, 34.69, 0.996, 0.996)$	$MAP_2(\mathcal{M}_1)$	$MAP_2(\mathcal{M}_2)$	$MMPP_2$
$\boldsymbol{\theta}^{(0)}$	$(0.148, 2.394, 0.069, 0.284)$	$(0.137, 2.336, 0.222, 0.863)$	$(0.423, 6.397, 0.374, 0.851)$
$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}\|\mathbf{t})$	$(0.552, 45.323, 0.007, 0.850)$	$(0.022, 36.65, 0.991, 0.996)$	$(0.241, 40.273, 0.276, 0.436)$
C_ω	$[0.464,\ 0.647]$	$[0.0183,\ 0.0238]$	$[0.231,\ 0.256]$
C_{λ_2}	$[40.371,\ 47.227]$	$[34.64,\ 38.61]$	$[34.68,\ 48.86]$
C_a	$[0.003,\ 0.218]$	$[0.985,\ 0.996]$	$[0.021,\ 0.598]$
C_b	$[0.831,\ 0.886]$	$[0.992,\ 0.997]$	$[0.336,\ 0.510]$
$C_{\mu_1}\ (0.746)$	$[0.711,\ 0.787]$	$[0.703,\ 0.791]$	$[0.736,\ 1.053]$
$C_{\sigma^2}\ (1.589)$	$[0.568,\ 1.780]$	$[1.37,\ 1.754]$	$[1.336,\ 1.824]$
$C_{R_1}\ (-0.320)$	$[0.002,\ 0.014]$	$[-0.325,\ -0.314]$	$[0.003,\ 0.036]$
$C_{R_2}\ (0.305)$	$[10^{-7},\ 2.4 \times 10^{-5}]$	$[0.303,\ 0.317]$	$[10^{-6},\ 0.006]$
$C_{R_3}\ (-0.318)$	$[0,\ 4.3 \times 10^{-7}]$	$[-0.315,\ -0.302]$	$[10^{-8},\ 0.0012]$
$\log f(\mathbf{t}\|model)$	-1000.66	818.81	-1112.11

Table 3: Performance of the MCMC for the simulated data set S3, under both types of MAP_2 and $MMPP_2$ (second to fourth columns, respectively). From top to bottom: starting values of the MCMC, posterior mean for the model parameters, 95% credible intervals for the model parameters, 95% credible intervals for the mean, variance, and first three autocorrelation coefficients of the inter-event times distribution, and marginal likelihoods.
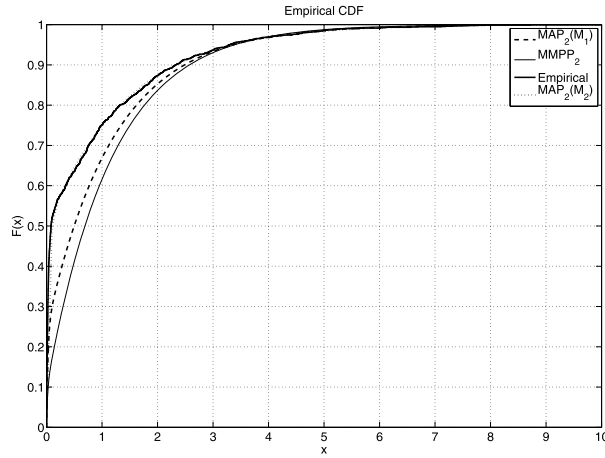
Figure 4: Empirical (thick solid line), and posterior predictive CDFs of the inter-event times via the $MAP_2(\mathcal{M}_1)$ (dashed line), the $MAP_2(\mathcal{M}_2)$ (dotted line) and the $MMPP_2$ (thin solid line) for the simulated data set S3.
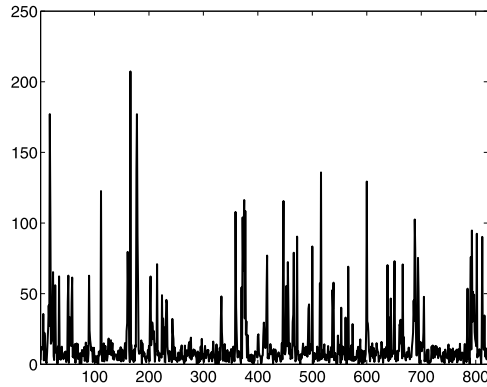

Figure 5: A total of 823 consecutive times (in sec) between real crashes.

period as in the simulated data case are considered. Figure 6 depicts the convergence of the algorithm by showing the evolution of the average parameters for the real data set.

Note from Figures 1 and 6 that none of the figures or Monte Carlo computations would change meaningfully if only a few thousand draws had been taken.

Table 4 and Figure 7 show in analogous way as in the simulated data, the performance of the MCMC algorithm. From the results it is clear that the MAP_2 of type $\mathcal{M}_1$ and the $MMPP_2$ outperform the MAP_2 of type $\mathcal{M}_2$, which presents the lowest marginal likelihood. Also, from Figure 7 it can be seen how the estimated CDFs under the MAP_2

Data set "Crashes logs"	$MAP_2(\mathcal{M}_1)$	$MAP_2(\mathcal{M}_2)$	$MMPP_2$
$\boldsymbol{\theta}^{(0)}$	$(0.062, 0.072, 0.6, 0.93)$	$(0.027, 0.087, 0.292, 0.359)$	$(0.102, 7.22, 0.639, 0.782)$
$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{t})$	$(0.286, 0.1102, 0.754, 0.936)$	$(0.574, 0.0805, 0.1838, 0.108)$	$(0.639, 0.138, 0.749, 0.832)$
C_ω	$[0.241, \ 0.337]$	$[0.401, \ 0.770]$	$[0.194, \ 0.875]$
C_{λ_2}	$[0.1005, \ 0.1208]$	$[0.075, \ 0.0867]$	$[0.125, \ 0.151]$
C_a	$[0.656, \ 0.837]$	$[0.013, \ 0.4749]$	$[0.155, \ 0.997]$
C_b	$[0.913, \ 0.956]$	$[0.069, \ 0.158]$	$[0.268, \ 0.992]$
$C_{\mu_1} \ (14.51)$	$[12.94, \ 16.45]$	$[13.63, \ 15.53]$	$[7.57, \ 35.06]$
$C_{\sigma^2} \ (496.58)$	$[297.6, \ 541.7]$	$[200.9, \ 310.9]$	$[78.9, \ 338.5]$
$C_{R_1} \ (0.328)$	$[0.1581, \ 0.299]$	$[-0.129, \ -4 \times 10^{-2}]$	$[0.15, \ 0.265]$
$C_{R_2} \ (0.203)$	$[0.1103, \ 0.245]$	$[9 \times 10^{-8}, \ 0.003]$	$[5 \times 10^{-4}, \ 0.202]$
$C_{R_3} \ (0.121)$	$[0.09, \ 0.165]$	$[-3 \times 10^{-4}, \ -10^{-7}]$	$[10^{-7}, \ 0.174]$
$\log f(\mathbf{t}|model)$	-2928.7	-3236.6	-2945.4

Table 4: Performance of the MCMC for the real data set, under both types of MAP$_2$ and MMPP$_2$ (second to fourth columns, respectively). From top to bottom rows: starting values of the MCMC, posterior mean for the model parameters, 95% credible intervals for the model parameters, 95% credible intervals for the mean, variance, and first three autocorrelation coefficients of the inter-event times distribution, and marginal likelihoods.
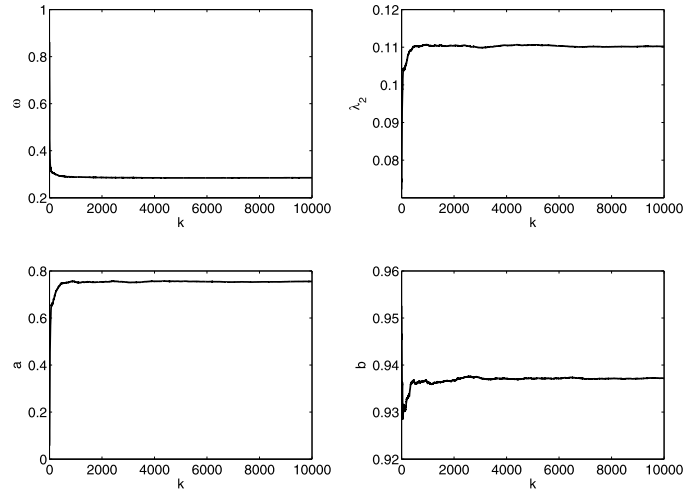
Figure 6: Average ω, λ_2 (top), a and b (bottom) from the first 10000 runs of MCMC for the real data set.
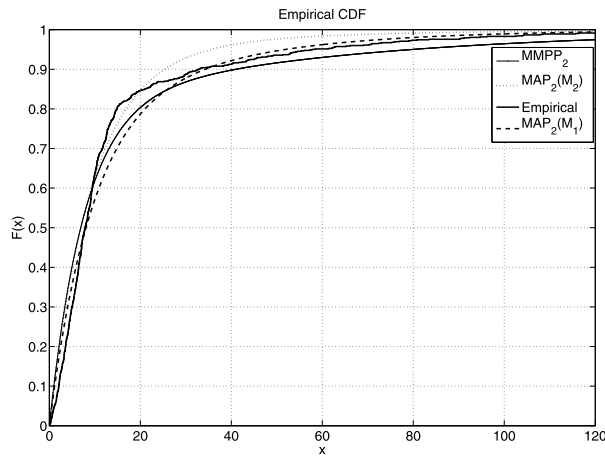

Figure 7: Empirical (thick solid line), and posterior predictive CDFs of the times between real crashes in a computer via the $MAP_2(\mathcal{M}_1)$ (dashed line), the $MAP_2(\mathcal{M}_2)$ (dotted line) and the $MMPP_2$ (thin solid line).

of type $\mathcal{M}_1$ and the $MMPP_2$ are closer to the empirical distribution than that under the MAP_2 of type $\mathcal{M}_2$. In all cases, the mean of the data are correctly fitted, while the $MMPP_2$ and MAP_2 of type $\mathcal{M}_2$ fail in capturing the sample variance. Both the MAP_2 of type $\mathcal{M}_1$ and the $MMPP_2$ underestimate the first autocorrelation coefficient but correctly fit the second and third coefficients. The MAP_2 of type $\mathcal{M}_2$ gives negative estimates for the first and third autocorrelation coefficients.
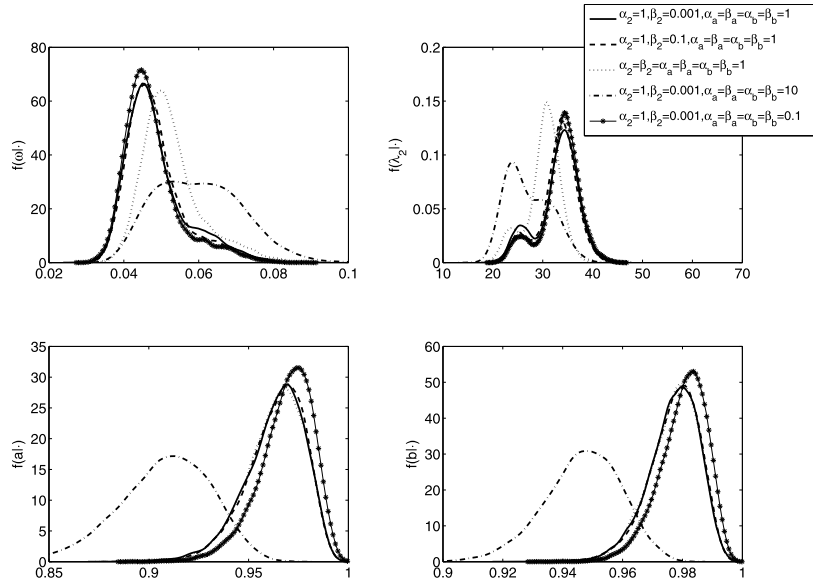
Figure 8: Posterior predictive distributions obtained when running the MCMC (MAP_2 of type $\mathcal{M}_1$) to the simulated data S1, under different choices of the prior structure.

3.5   Prior sensitivity

In this section we aim to test the prior sensitivity of the proposed MCMC algorithm. As seen just before Section 3.1, the prior structure was set as follows

$$
\begin{aligned}
\omega &\sim \mathcal{U}(0,1), \\
\lambda_2 &\sim Ga\left(\alpha_2, \beta_2\right), \\
a &\sim Be(\alpha_a, \beta_a), \\
b &\sim Be(\alpha_b, \beta_b),
\end{aligned}
$$

where in the numerical experiments carried out in Section 3.4, the values $\alpha_2 = 1$, $\beta_2 = 0.001$, and $\alpha_a = \alpha_b = \beta_a = \beta_b = 1$ were selected. Figure 8 depicts in solid line the estimated (posterior) densities for the model parameters obtained when running the MCMC (MAP_2 of type $\mathcal{M}_1$) for the S1 data set, under the previous prior structure. Note that the summarized information of such posterior sample is given in the second column of Table 1. Also, Figure 8 represents the posterior distribution of the parameters of the MAP_2 (type $\mathcal{M}_1$), under different choices of the priors' distributions. In particular, in dashed and dotted lines the densities under the same priors for a and b as in the solid line case is shown, while the distribution of λ_2 becomes more informative (the variance decreases from 10000 to 100 and 0.1 in the dashed and dotted cases, respectively). The posterior densities shown in dashed–dotted and starred solid line are obtained under the same prior structure for λ_2 as in the solid line case, while the priors for parameters a and b change to a more (respectively less) informative distributions. Only in the case

where $\alpha_a = \alpha_b = \beta_a = \beta_b = 10$, the posteriors differ significantly from those obtained in the rest of the cases. However, this is a very large change in the prior distribution and we should expect this to be influential when we have relatively small data samples as is the case here. Since analogous results were obtained when testing the prior sensitivity for the rest of data bases, it can be concluded that there is little sensitivity in posterior inference as long as the prior densities for a and b are relatively uninformative.

4   Inference for the MAP$_2$/G/1 queueing system

In this section, we shall consider the MAP_2 as a model for the arrival process in a single-server, first in first out queueing system with independent, general service times. Following standard notation, we shall denote this queueing system by $MAP_2/G/1$. The inference approach described in Section 3 will be combined with techniques from the queueing literature in order to estimate predictive equilibrium distributions for this system.

4.1   The MAP$_2$/G/1 queueing system

The main properties of the $MAP_2/G/1$ queueing system are briefly outlined below. For the complete description of the queueing system we refer the reader to Ramaswami (1980), Lucantoni (1993) or Ramírez-Cobo et al. (2014a).

Assume that we have a queueing system with MAP_2 arrivals, specified by the rate matrices $\{D_0, D_1\}$ and independent and identically distributed service times with a general distribution, which are independent of the arrival times. Denote by $\mu^\star < \infty$ the expected value of the service time. Then, the traffic intensity of this system is given by

$$\rho = \lambda^\star/\mu^\star, \tag{22}$$

where $\lambda^\star$ is the stationary arrival rate (inverse of the expected inter-event time), defined as

$$\lambda^\star = \boldsymbol{\pi} D_1 \mathbf{e}^T = 1/\mu_1,$$

where μ_1 is defined as in Eq. (3).

Now define $Z(t)$ to be the number of customers in the system (including in service, if any) at time t and let τ_k be the epoch of the k-th departure from the queue, with $\tau_0 = 0$. If the system is stable ($\rho < 1$), then the stationary distributions of the considered queueing system are defined as follows. For $i \geq 1$, define

$$z_i = \lim_{k \to \infty} P\left[Z(\tau_k) = i\right],$$

which represents the stationary probability that the queue length is equal to i when a departure occurs. Similarly, denote by y_i the stationary probability that the queue length is equal to i, at an arbitrary time. Finally, consider the waiting time CDF denoted by $W(x)$ and defined as the probability that at an arbitrary time, a virtual customer who arrives at that time waits at most a time x before entering service. Closed-form expressions for the generating functions of the queue length distributions and for the

Laplace Stieltjes transform of the waiting time distribution can be found in Lucantoni (1993). In order to invert such transforms the numerical routines described in Lucantoni (1993), as well as in Abate and Whitt (1995) were implemented.

4.2   Bayesian estimation of the $\mathrm{MAP}_2/G/1$ queueing system

As commented in Section 1, in some real-life contexts as teletraffic or insurance it is important to estimate the steady-state distributions of queueing systems. In particular, since the MAP_2 allows for dependent observations, the inference for the $MAP_2/G/1$ queue is of interest. However, inferential analysis of the $MAP/G/1$ queueing system are rare. As established by Mcgrath et al. (1987) and Armero and Bayarri (1999), the Bayesian reasoning to the theory of queues presents a number of advantages with respect to other approaches as that related to the restrictions of the parameter space, accuracy of estimators, prediction or transient analyses. In this section we combine the MCMC algorithm shown in Section 3.2 with theoretical results from Section 4.1 to obtain estimations of the queue length and waiting time distributions in a $MAP_2/G/1$ queueing system.

Given a sample of inter-event data, we have seen that the Gibbs sampler can be used to produce a sample of values $\boldsymbol{\theta}^{(k)} = (\omega^{(k)}, \lambda_2^{(k)}, a^{(k)}, b^{(k)})$, for $k = 1, \ldots, M$, from the posterior distribution of the MAP_2 parameters. Assuming that the service rate $\mu^\star$ is known, it is straightforward to estimate the probability that the system is stable,

$$P(\rho < 1 | data) \approx \frac{1}{M} \sum_{k=1}^{M} I\left(\rho^{(k)} < 1\right), \qquad (23)$$

where $\rho^{(k)}$ is the system traffic intensity calculated from Eq. (22) using $\boldsymbol{\theta}^{(k)}$ and $I(\cdot)$ is the indicator function. If this probability is high, then for each set $\boldsymbol{\theta}^{(k)}$ of generated parameters such that $\rho^{(k)} < 1$, the conditional posterior distributions of queue length and waiting times, given stability, can be estimated by Rao Blackwellization (Blackwell, 1947), that is, by simply averaging over the parameters satisfying the stability condition. Thus, for example, the predictive distribution of the waiting time $W(x)$, is estimated by

$$\frac{1}{C} \sum_{c=1}^{C} W^{(c)}(x), \qquad (24)$$

where $W^{(c)}(x)$ results from inverting the Laplace–Stieltjes transform of the stationary time distribution after substituting the set of parameters satisfying the stability condition $\boldsymbol{\theta}^{(c)}$, $c = 1, \ldots, C \leq M$.

Note finally that when the service parameter $\mu^\star$ is unknown, then, given an independent sample of service time data, inference for the service rate can be carried out along the lines of, for example, Armero and Bayarri (1994) (for the case where the service times were exponentially distributed). For a Monte Carlo sample, $\mu^{\star(1)}, \ldots, \mu^{\star(M)}$ from the posterior distribution of the service rate, the traffic intensity may be estimated by calculating $\rho^{(k)}$ given $(\boldsymbol{\theta}^{(k)}, \mu^{\star(k)})$ and averaging as in (23). In order to condition on the

existence of equilibrium, only those parameter sets $(\boldsymbol{\theta}^{(k)}, \mu^{\star(k)})$ such that $\rho^{(k)} < 1$ are retained.

4.3 Application to Internet traffic analysis

Here we estimate the probabilities of congestion, queue lengths and waiting time distributions in a $MAP_2/G/1$ queue under a sample of real arrival data. The analyzed data set represents a sequence of 300 consecutive inter-arrival times (in sec) and can be found in the *Internet Traffic Archive* (BC-pAug89 trace), `http://ita.ee.lbl.gov/html/traces.html`. The same data set was explored in Ramirez et al. (2008); Ramírez-Cobo et al. (2010a) where Bayesian estimation of queueing systems with heavy-tailed arrival processes was considered. However, in such works the empirical autocorrelation of the inter-arrival times was not taken into account. The trace of the inter-arrival times as well as some of its statistical descriptors (sample mean, variance and first three autocorrelations coefficients) is shown by the top panel of Figure 9.

Both types of MAP_2 as well as the $MMPP_2$ were fitted to the data. In this case, the MAP_2 of type $\mathcal{M}_1$ showed the best performance. The bottom panel of Figure 9 shows the empirical CDF of the inter-arrival times and superimposed (in dashed line) the fitted MAP_2 distribution generated using the Gibbs sampler described in Section 3.2. Also, the figure shows the 95% credible intervals for the statistical descriptors given by the top panel of the same figure.

Now we shall consider the queueing aspects. Given the MAP_2 arrival process, for computational simplicity we shall assume an exponential service process with rate $\mu^\star$. For the considered real dataset (of length equal to 300) and under the described setting, the computational time for estimating the queue according to Section 4.2 was around 32 seconds per 1000 iterations. Table 5 shows the posterior probability of stability (third column) and the expected value for the traffic intensity (fourth column) for an assortment of values of $\mu^\star$ (the expected service time is $E(S) = 1/\mu^\star$). As expected, when $\mu^\star$ is large (faster service on average), the probability of stability of the system increases. From the table, it is clear that there is a high probability that the system is stable (that is, no congestion occurs) for values of $\mu^\star$ greater than 450.

In Figures 8 and 9, we illustrate the predictive steady-state queue length (at departures and at an arbitrary time) and waiting time distributions for some values of $\mu^\star$ greater than 450. Figure 10 depicts the predictive queue length distributions at departures and at an arbitrary time. As would be expected, for faster services on average (large $\mu^\star$) the probability of an empty system is larger than for slower services. Finally, Figure 11 shows the predictive waiting time distributions for the values of $\mu^\star$ considered for the analysis. It can be seen that when the service rate increases, then the distribution of the waiting times decreases, as expected as well.

5 Conclusions

In this article, we have illustrated how to carry out Bayesian inference for the stationary MAP_2 and then combined this approach with results from the queueing theory
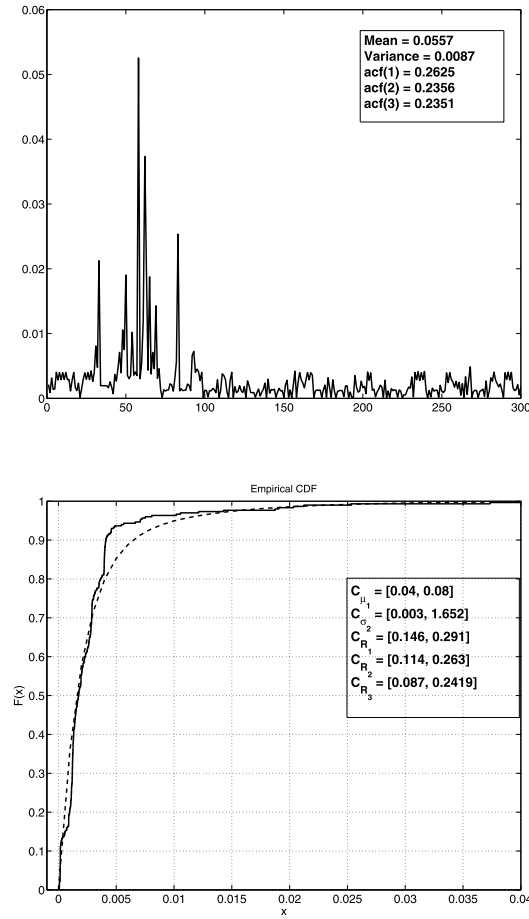
Figure 9: Top panel: trace of the teletraffic inter-arrival times and sample statistical descriptors (mean, variance and first three autocorrelation coefficients). Bottom panel: empirical CDF of the inter-arrival times (solid line), fitted distribution by the MAP_2 (dashed line) and 95% credible intervals for the statistical descriptors.

to estimate the steady-state distributions of interest in a $MAP_2/G/1$. The MAP_2 has proven suitable for many statistical modeling applications since it combines phase-type distributions with a specific autocorrelation function for the inter-event times, allowing in this way for the modeling of dependent observations. In this work, the problem of the non-identifiability of the MAP_2 has been overcome by using the canonical version of the process instead of the redundant form. Extensive comparisons with the $MMPP_2$, a simplified MAP_2 widely studied in the literature, have also been provided. A Gibbs sampler has been combined with a Chib criteria for selection model, although a different approach as the reversible jump MCMC (Green, 1995) could have been used instead. The result is an efficient algorithm although the computational cost may be notable,

$\mu^\star$	$E(S)$	$\mathbb{P}(\rho < 1|data)$	$\mathbb{E}(\rho|data)$
1500	0.0006	1	0.2379
1000	0.001	1	0.3569
500	0.002	0.9941	0.7138
450	0.0025	0.9436	0.7931
400	0.0025	0.7605	0.8923
395	0.002531	0.7330	0.9036
390	0.002564	0.7043	0.9152
385	0.002597	0.6752	0.9271
375	0.002667	0.6136	0.9518
350	0.002285	0.4594	1.0198
325	0.0030768	0.3138	1.0982
300	0.003334	0.2167	1.1921

Table 5: Expected service time (second column), probability that the system is stable (third column) and traffic intensity (fourth column).
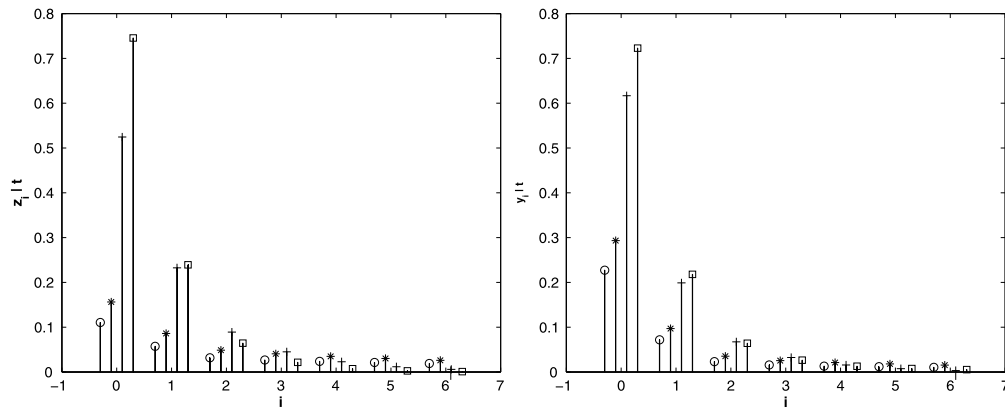
Figure 10: Predictive queue length distribution at departures times and at an arbitrary time for the Internet data for an assortment of service rates ($\circ : \mu^\star = 450$, $* : \mu^\star = 500$, $\diamond : \mu^\star = 700$, $+ : \mu^\star = 1000$, $\square : \mu^\star = 1500$).

especially when considering the inference for the $MAP_2/G/1$ system due to the need of inverting a set of generating functions at each iteration.

A number of extensions are possible, among them the estimation of higher state MAPs, which are expected to show more versatility for modeling purposes than the MAP_2, and of the batch Markov arrival process, the $BMAP$ (Lucantoni, 1993), which allows for correlated arrivals in batches. Generalizations to inference for the $BMAP/G/1$ queue are also of interest. We are aware of both the theoretical and computational complexities of such problems due to the lack of unique representations and the increasing number of parameters. These complications present a challenging problem that we hope to address in the future.
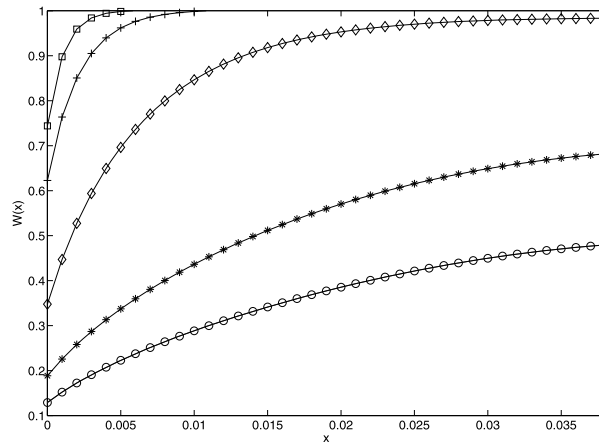
Figure 11: Predictive waiting time distribution for the Internet data for an assortment of service rates ($\circ : \mu^\star = 450$, $* : \mu^\star = 500$, $\diamond : \mu^\star = 700$, $+ : \mu^\star = 1000$, $\square : \mu^\star = 1500$).

References

Abate, J. and Whitt, W. (1995). "Numerical Inversion of Laplace transforms of probability distributions." *ORSA Journal on Computing*, 7(1): 36–43.   1186

Armero, C. and Bayarri, M. (1994). "Bayesian prediction in $M/M/1$ queues." *Queueing Systems*, 15: 401–417. MR1266803. doi: http://dx.doi.org/10.1007/BF01189248. 1186

Armero, C. and Bayarri, M. (1999). "Dealing with uncertainties in queues and networks of queues: a Bayesian approach." *Multivariate Analysis, Design of Experiments and Survey Sampling*. MR1722512.   1186

Asmussen, S. and Albrecher, H. (2010). *Ruin probabilities. Advanced Series on Statistical Science & Applied Probability*. World Scientific. MR2766220. doi: http://dx.doi.org/10.1142/9789814282536.   1164

Asmussen, S. and Koole, G. (1993). "Marked point processes as limits of Markovian arrival streams." *Journal of Applied Probability*, 30: 365–372. MR1212668.   1168

Banerjee, A., Gupta, U., Horváth, G., and Chakravarthy, S. (2015). "Analysis of a finite-buffer bulk-service queue under Markovian arrival process with batch-size-dependent service." *Computers & Operations Research*, 60: 138–149. MR3328044. doi: http://dx.doi.org/10.1016/j.cor.2015.02.012.   1163

Blackwell, D. (1947). "Conditional expectation and unbiased sequential estimation." *Annals of Mathematical Statistics*, 18: 105–110. MR0019903.   1186

Bodrog, L., Heindlb, A., Horváth, G., and Telek, M. (2008). "A Markovian canonical form of second-order matrix-exponential processes." *European Journal of Operational Research*, 190: 459–477. MR2412984. doi: http://dx.doi.org/10.1016/j.ejor.2007.06.020.   1164, 1166, 1168

Casale, G., Z. Zhang, E., and Simirni, E. (2010). "Trace data characterization and fitting for Markov modeling." *Performance Evaluation*, 67: 61–79. 1163

Çinlar, E. (1975). *Introduction to stochastic processes*. Usa: Prentice–Hall. MR0380912. 1171

Chakravarthy, S. (2001). "The Batch Markovian arrival process: a review and future work." In et al., A. K. (ed.), *Advances in probability and stochastic processes*, 21–49. 1165, 1166

Chaudhry, M., Singh, G., and Gupta, U. (2013). "A simple and complete computational analysis of $MAP/R/1$ queue using roots." *Methodology and Computing in Applied Probability*, 15: 563–582. MR3085880. doi: http://dx.doi.org/10.1007/s11009-011-9266-3. 1163

Cheung, E. and Landriault, D. (2010). "A generalized penalty function with the maximum surplus prior to ruin in a MAP risk model." *Insurance: Mathematics and Economics*, 46: 127–134. MR2586163. doi: http://dx.doi.org/10.1016/j.insmatheco.2009.07.009. 1163

Cheung, E. and Runhuan, F. (2013). "A unified analysis of claims costs up to ruin in a Markovian arrival risk model." *Insurance: Mathematics and Economics*, 53: 98–109. MR3081465. doi: http://dx.doi.org/10.1016/j.insmatheco.2013.04.001. 1163

Chib, S. (1995). "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90: 1313–1321. MR1379473. 1173

Dudina, O., Kim, C., and Dudin, S. (2013). "Retrial queueing system with Markovian arrival flow and phase-type service time distribution." *Computers & Industrial Engineering*, 66(2): 360–370. 1163

Fearnhead, P. and Sherlock, C. (2006). "An exact Gibbs sampler for the Markov-modulated Poisson process." *Journal of the Royal Statistical Society: Series B*, 65(5): 767–784. MR2301294. doi: http://dx.doi.org/10.1111/j.1467-9868.2006.00566.x. 1164

Green, P. (1995). "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82: 711–732. MR1380810. doi: http://dx.doi.org/10.1093/biomet/82.4.711. 1188

Gruet, M., Philippe, A., and Robert, C. (1999). "MCMC control spreadsheets for exponential mixture estimation." *Journal of Computational and Graphical Statistics*, 8: 298–317. 1169

Hervé, L. and Ledoux, J. (2013). "Geometric rho-mixing property of the inter-arrival times of a stationary Markovian Arrival Process." *Journal of Applied Probability*, 50: 598–601. MR3102503. doi: http://dx.doi.org/10.1239/jap/1371648964. 1166

Kang, S. and Sung, D. (1995). "Two-state MMPP modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times." In *Global*

Telecommunications Conference, 1995, GLOBECOM'95, IEEE, volume 2, 1422–
1426. IEEE.   1168, 1169

Kriege, J. and Buchholz, P. (2010). "An empirical comparison of MAP fitting al-
gorithms." In *International GI/ITG Conference on Measurement, Modelling, and
Evaluation of Computing Systems and Dependability and Fault Tolerance*, 259–273.
Springer.   1164

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods
in stochastic modeling*, volume 5. SIAM. MR1674122. doi: http://dx.doi.org/
10.1137/1.9780898719734.   1165

Lucantoni, D. (1991). "New results for the single server queue with a Batch Markovian
Arrival Process." *Stochastic Models*, 7: 1–46. MR1102528. doi: http://dx.doi.org/
10.1080/15326349108807174.   1163

Lucantoni, D. (1993). "The $BMAP/G/1$ queue: A tutorial." In Donatiello, L. and
Nelson, R. (eds.), *Models and Techniques for Performance Evaluation of Computer
and Communication Systems*, 330–358. New York: Springer.   1163, 1165, 1185, 1186,
1189

Mcgrath, M., Gross, D., and Singpurwalla, N. (1987). "A subjective Bayesian approach
to the theory of queues I-Modeling." *Queueing Systems*, 1(4): 317–333. MR0899679.
doi: http://dx.doi.org/10.1007/BF01150668.   1186

Montoro-Cazorla, D. and Pérez-Ocón, R. (2014). "A reliability system under differ-
ent types of shock governed by a Markovian arrival process and maintenance pol-
icy K." *European Journal of Operational Research*, 235(3): 636–642. MR3166158.
doi: http://dx.doi.org/10.1016/j.ejor.2014.01.021.   1163

Neuts, M. F. (1974). *Probability distributions of phase type*. Purdue University. Depart-
ment of Statistics.   1163

Neuts, M. F. (1979). "A versatile Markovian point process." *Journal of Applied Proba-
bility*, 16: 764–779. MR0549556.   1163

Okamura, H., Dohi, T., and Trivedi, K. (2009). "Markovian arrival process parameter
estimation with group data." *IEEE/ACM Transactions on Networking*, 17(4): 1326–
1339.   1163

Prabhu, N. (1998). *Stochastic storage processes: queues, insurance risk, dams,
and data communication*. Springer Science & Business Media. MR1492990.
doi: http://dx.doi.org/10.1007/978-1-4612-1742-8.   1164

Ramaswami, V. (1980). "The N/G/1 queue and its detailed analysis." *Advances in Ap-
plied Probability*, 222–261. MR0552956. doi: http://dx.doi.org/10.2307/1426503.
1163, 1185

Ramirez, P., Lillo, R., and Wiper, M. (2008). "Bayesian analysis of a queueing sys-
tem with a long-tailed arrival process." *Communications in Statistics – Simulation
and Computation*, 37(4): 697–712. MR2744052. doi: http://dx.doi.org/10.1080/
03610910701753861.   1187

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010a). "Bayesian inference for double Pareto lognormal queues." *The Annals of Applied Statistics*, 4(3): 1533–1557. MR2758340. doi: http://dx.doi.org/10.1214/10-AOAS336. 1164, 1187

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010b). "Nonidentifiability of the two-state Markovian arrival process." *Journal of Applied Probability*, 47(3): 630–649. MR2731339. doi: http://dx.doi.org/10.1239/jap/1285335400. 1164, 1166, 1168, 1174

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2014a). "Identifiability of the MAP2/G/1 queueing system." *Top*, 22(1): 274–289. MR3190824. doi: http://dx.doi.org/10.1007/s11750-012-0254-8. 1185

Ramírez-Cobo, P., Marzo, X., Olivares-Nadal, A., Francoso, J., Carrizosa, E., and Pita, M. (2014b). "The Markovian arrival process: a statistical model for daily precipitation amounts." *Journal of Hydrology*, 510: 459–471. 1163

Riska, A., Squillante, M., Yu, S.-Z., Liu, Z., and Zhang, L. (2002). "Matrix-analytic analysis of a MAP/PH/1 queue fitted to web server data." *Matrix-Analytic Methods; Theory and Applications*, 333–356. MR1923893. 1164

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2015). "Failure modeling of an electrical N-component framework by the non-stationary Markovian arrival process." *Reliability Engineering & System Safety*, 134: 126–133. 1163

Rydén, T. (1996). "On identifiability and order of continous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions." *Journal of Applied Probability*, 33: 640–653. MR1401462. 1166

Scott, S. (1999). "Bayesian analysis of a two-state Markov modulated Poisson process." *Journal of Computational and Graphical Statistics*, 8(3): 662–670. 1164

Scott, S. (2002). "Bayesian methods for hidden Markov models. Recursive Computing in the 21st Century." *Journal of the American Statistical Association*, 457: 337–351. MR1963393. doi: http://dx.doi.org/10.1198/016214502753479464. 1171

Scott, S. and Smyth, P. (2003). "The Markov Modulated Poisson Process and Markov Poisson Cascade with applications to web traffic modeling." *Bayesian Statistics*, 7: 1–10. MR2003531. 1164

Telek, M. and Horváth, G. (2007). "A minimal representation of Markov arrival processes and a moments matching method." *Performance Evaluation*, 64(9): 1153–1168. 1164, 1166

Wu, J., Liu, Z., and Yang, G. (2011). "Analysis of the finite source MAP/PH/N retrial G-queue operating in a random environment." *Applied Mathematical Modelling*, 35(3): 1184–1193. MR2740860. doi: http://dx.doi.org/10.1016/j.apm.2010.08.006. 1163

Xue, J. and Alfa, A. (2011). "Geometric tail of queue length of low-priority customers in a nonpreemptive priority MAP/PH/1 queue." *Queueing Systems*, 69(1): 45–76. MR2835230. doi: http://dx.doi.org/10.1007/s11134-011-9221-6. 1163

Zhang, M. and Hou, Z. (2011). "Performance analysis of MAP/G/1 queue with working vacations and vacation interruption." *Applied Mathematical Modelling*, 35(4): 1551–1560. MR2763797. doi: http://dx.doi.org/10.1016/j.apm.2010.09.031. 1163

Acknowledgments